Alpha vs. Omega in Investment Performance

Christopher Lewis | June 2022

Read Time: 30 mins

There are some games you feel like you're in control of the game, but you're not in control of the score. That's kind of the way I felt in the Super Bowl. We moved the ball. We were able to stop them on third down. I didn't feel like we were being dominated on the field. But we were being dominated on the scoreboard.

– Football coach Bill Belichick, recounting his real-time performance evaluation while trailing 28-3 in the second half of Super Bowl LI[1].

Summary

A typical goal of actively managed investment products is to achieve higher returns relative to some benchmark, such as an index or peer group. In a competitive market higher returns are difficult to achieve, and even harder to prove as coming from repeatable skill rather than temporary luck. The standard method for identifying such skill is to study past returns that a manager (or model) has produced. However, much like a football scoreboard, relative returns can diverge from an underlying skill advantage for frustratingly long time horizons. To combat the noisiness of returns, introducing additional performance measures such as "Omega" can provide a more diverse perspective of performance. The concept of Omega is built upon simple, binary measurements of winners vs. losers in a portfolio, and how the proportion of individual victories compares to the applicable benchmark.

Key Takeaways

1. Past returns can be a fickle scoreboard for predicting future outperformance (Alpha).
   - The signal-to-noise ratio in investment returns is very low, meaning it takes a *long* track record of past returns to reach any scientific level of confidence that skill exists.
2. Including other performance metrics, such as Omega, can add a diversifying perspective to returns-based measures, and unearth other insights about manager (or model) behavior.
3. Omega builds upon simple, binary measurements of winners vs. losers in a portfolio.
   - In the long-run Alpha and Omega performance is correlated, but not perfectly so, which provides the "diversification" benefit of different vantage points.

[1] For those unfamiliar, Belichick's team famously scored the next 31 points and won the game.

Return-Based Metrics (Alpha)

Importance of Returns

To state the obvious, past returns are important to study because they represent the goal for the future. At any point in time, for whatever goals an investor may have, the ability and timetable to reach those goals will depend on the future returns produced by their investments, which themselves will be aided or hindered by the relative returns (net of fees) from any active management decisions.

The benefit of even modest levels of return outperformance can be significant. Consider Table 1 below, which summarizes the growth of one dollar invested at various levels of outperformance on top of the U.S. equity market's historical compound return rate of ~10%.

Table 1. Future Value of $1.00 Invested

For illustrative purposes only; the following example does not pertain to any investable product.

Years	Annual Outperformance above 10.0% Market Return				
Invested	0.0%	0.5%	1.0%	1.5%	2.0%
5	$1.61	$1.65	$1.69	$1.72	$1.76
10	2.59	2.71	2.84	2.97	3.11
15	4.18	4.47	4.78	5.12	5.47
20	6.73	7.37	8.06	8.82	9.65
25	10.83	12.14	13.59	15.20	17.00

Source: Ken French data library and WEDGE Capital Management. 10% market return calculated as the cap-weighted, geometric average return from July 1926 – March 2022 for all CRSP firms incorporated in the U.S. and listed on the NYSE, AMEX, or NASDAQ.

Difficulties of Returns

Unfortunately, past returns bring with them a litany of complications. These issues are not insurmountable, but require long, uninterrupted time horizons to even potentially see through them – longer than what many managers possess as a track record. Comparison indices often have heavy market cap skewness in their weights, and periodically experience large outlier returns (e.g. GameStop in January 2021), both of which exert a gravitational pull on the overall benchmark's return. In addition, macroeconomic shocks like recessions, wars, pandemics, etc., as well as investor taste cycles of "irrational exuberance" (e.g. the Dot-Com Bubble) can be unpredictable and unavoidable headwinds for even a skilled manager. All these items affect the following summary description of live investment returns by raising the amount of statistical "noise" that is beyond a manager's control:

$$Realized\ Returns = Expected\ Returns \pm Noise$$

As any strategy rolls through live periods, its realized returns are some combination of expected returns ("expected" referring to the mathematical operation, not a single prediction; see the appendix for further discussion) and the "unknown unknowns" which are impossible to predict but inevitably occur.

Compared to other scientific endeavors, investment returns contain a great deal of noise for a small amount of signal. To demonstrate this point, consider a theoretical manager with *guaranteed* skill of 2% average annual outperformance and a 4% tracking error. By assuming their skill, it is an absolute certainty[2] that given enough time this manager will outperform and earn their fees. The tracking error creates short-term variability though, such that the manager can temporarily underperform.

Just how long that temporary underperformance can persist may come as a surprise. Table 2 summarizes one million simulated track records for this manager across various investment horizons, and how likely different levels of trailing relative returns are to occur. In Fixed Income investing this type of analysis is often referred to as a "Value at Risk" assessment.

Table 2. Value at Risk Simulations for a Skillful Manager

Average annual relative returns (in %s), simulated from quarterly relative returns with mean annual outperformance of 2% and tracking error of 4%.

Invest	Probability of Occurrence				
Horizon	1%	5%	10%	25%	50%
3Y	-3.37	-1.80	-0.97	0.44	2.00
5Y	-2.17	-0.94	-0.29	0.79	2.00
7Y	-1.52	-0.49	0.06	0.98	2.00
10Y	-0.95	-0.08	0.38	1.15	2.00
15Y	-0.41	0.30	0.68	1.30	2.00
20Y	-0.07	0.53	0.85	1.40	2.00
25Y	0.14	0.68	0.98	1.46	2.00

Source: WEDGE Capital Management. This illustrative example does not pertain to any investable product.

Note: Table 2 is generated by randomly sampling a Normal Distribution with mean and standard deviation equal to the stipulated quarterly outperformance and tracking error. The assumption of Normality will not hold for many managers, although that is often due to wider tails in practice than exist in a Normal Distribution. In such cases the same analysis could be generated from bootstrap samples of empirical data, which for wider tails would indicate even more extreme relative downturns than are shown above. Either type of simulation is only possible when the time series of relative returns is independent across periods, which can be tested via ARMA models, and tends to be true over short time measurement horizons.

To unpack Table 2, first look to the far-right column. In half of the simulated performance paths the skillful manager's trailing outperformance is 2% per year or less. That finding should come as no surprise: it matches the average outperformance assumed into the simulations. Columns to the left are

[2] Of course, in practice no such guarantees can be made, which is what makes this thought experiment so compelling. Even working within the known Normal Distribution, the odds and duration of temporary underperformance are far from trivial.

more interesting, as they show relative drawdown levels based on their probability of occurrence. Over a 10-year investment horizon, even a manager with known[3] skill of 2% per year has a 5% chance of underperforming their benchmark. Or, to use an insurance analogy, a "20-Year Storm" in the markets can ruin a 10-year track record for even a skillful investor. The 5% probability column is particularly interesting given the prominent use in statistics of 5% confidence thresholds to "prove" that an observed result was not just a lucky sample. The FDA, for example, relies on that 5% confidence standard for drug approval decisions (Kennedy-Shafer, 2017). Imagine if drug trials were still giving false negatives after ten years of patient trials – that's the magnitude of noise inherent in investment returns!

Academic Perspectives

The realization that even multi-year return horizons fail to conclusively reveal manager skill is not novel. See Carhart (1997), Arnott et al. (2017), Cornell et al. (2017), and Harvey and Liu (2018) for additional tests on past manager returns. These studies generally agree that using five years or less of trailing returns is a futile exercise at best, and in some cases damaging (i.e. mean reversion effects are powerful enough that one may be better off replacing recent outperformers with recent underperformers).

At the same time, it is often impractical to study the truly long-term data, because (1) people retire and models may be changed over time; and (2) the problem of survivorship bias in manager databases expands as the time horizon grows. Instead of waiting for 20+ years to roll by while maintaining a fair, unbiased database, a different approach is to supplement the due diligence toolkit with additional measures of performance. Some of the cited researchers above, who reject the usefulness of short-term past returns, suggest other data points as better indicators for the future. Arnott et al. (2017), for example, use valuation measures to delineate success caused by potentially fleeting multiple expansion versus improving fundamentals. Cremers and Pareek (2016) suggest active share and turnover are key ingredients for persistent Alpha. WEDGE Capital supports both of those recommendations as useful information to consider, but places even more importance on a different perspective altogether: that of binary win/loss outcomes at the individual holdings level. This dimension of performance is known internally within WEDGE as "Omega."

Holdings-Based Metrics (Omega)

Performance metrics that drill down to a manager's individual holdings gain a statistical benefit of having more data points to study[4], but exacerbate the problem of noisiness. At least for stocks, individual ex-post outcomes offer no useful insights on their own, so some type of aggregation is still

[3] To stress again, this level of certainty does not exist in practice. Manager skill can only be estimated, it cannot be known.

[4] Namely, that the Law of Large Numbers should provide better clarity in statistical tests for skill.

necessary. Portfolio returns accomplish that aggregation by summing up the individual position-weighted returns at each point in time.

Portfolio Batting Average

For a simple alternative to returns that reaches towards that ideal of holdings-level data power, ignore position weights and condense each holding's return into a binary measure of whether it had a higher return than the benchmark or not over a particular period. That list of 1s and 0s can be aggregated to the portfolio level, creating Batting Averages of the percentage of holdings that outperformed the benchmark in each period. Equation 1 explicitly defines the Batting Average for a N stock portfolio during period t. Each period's Batting Average will always fall between 0% and 100%, although for a diversified portfolio over short time horizons, will typically be drawn towards ~50% on average[5].

$$BA_t = \frac{1}{N}\sum_{i=1}^{N} win_{i,t}\,; \tag{1}$$

$$win_{i,t} = \begin{cases} 1 \ \text{ for } \ return_{i,t} > return_{bench,t} \\ 0 \ \text{ for } \ return_{i,t} \leq return_{bench,t} \end{cases}$$

Batting Average Advantage

The natural extension of Batting Average to relative performance is to subtract the Batting Average of the benchmark itself (i.e. the percentage of constituents with higher returns than the index) from that of the portfolio each period. This relative metric is called the Batting Average Advantage ("BAA"). Like relative returns, an active manager's goal is to produce positive BAA over time, ideally to such an extent that statistical tests on the average BAA across periods (Equation 2) conclude that there is some true, underlying skill in selecting stocks with higher returns into the portfolio. Long-term BAA averages tend to converge near 0%, on similar scales to Alpha measures (i.e. 1% – 2% BAA can be quite impactful).

$$\overline{BAA} = \frac{1}{T}\sum_{t=1}^{T} BA_{port,t} - BA_{bench,t} \tag{2}$$

BAA eliminates two sources of noise from returns: weight imbalances and outliers. If only a few stocks in a benchmark outperform in a given period (e.g. the heavily weighted mega-cap stocks perform well, or a few individual stocks post outrageously high returns), then the benchmark's Batting Average for that period will reflect the small pool of potential winners, as will the portfolio's Batting Average.

[5] For most equity indices, slightly less than half of its constituents outperform in a typical period. This mild asymmetry stems from (1) compounding effects and (2) limited liability protections of public stockholders. A 50% loss is more damaging to the growth of capital than a 50% gain is helpful to it, and stock returns cannot be worse than –100% but they can exceed +100%.

Extension to Omega

Batting Average Advantage can be standardized into a metric called Omega, which controls for portfolio size and turnover characteristics to compare estimated skill consistently across different managers and styles. This standardization is analogous to how the true definition of "Alpha" controls for beta exposure(s) in a returns-based performance analysis. In practice "Alpha" is colloquially overused to refer to any type of returns outperformance, rather than its formal definition of a regression intercept. Omega's relationship with Batting Average Advantage is no different. If the scope of a performance analysis is limited to similar managers or investment approaches, Batting Average Advantage and Omega can be thought of as synonyms. The following section leans on this fluid parlance, to demonstrate some of the practical insights that BAA can reveal on its own. The appendix provides a formal definition of Omega and describes what technically distinguishes it from BAA.

Omega in Practice

Improved Perspective

Viewing both the Alpha and Omega dimensions of performance can add an extra depth of perspective to performance studies. For a long-term analysis – such as deciding between backtested models, for example – seeing balanced outperformance on both relative returns and BAA is a reassuring sign of stability, and that "data snooping" concerns are less likely to be present. For reviewing shorter-term, realized results, the imbalance of Alpha and Omega can be a helpful way to quickly identify periods of high noise in returns.

Figure 1 on the following page provides a succinct example of both the long-term and short-term benefits to including Omega in the thought process. The scatter plot shows live, quarterly returns for an anonymous[6] domestic equity product with a sufficiently long history of returns and holdings data (herein referred to as "Portfolio X"). Each historical quarter's Batting Average Advantage is marked on the x-axis, with relative returns for that same quarter marked on the y-axis. The long-term relationship is easily apparent: higher BAAs tend to correspond with higher return outperformance. The correlation across all points is 48%, shown as a dashed blue line.

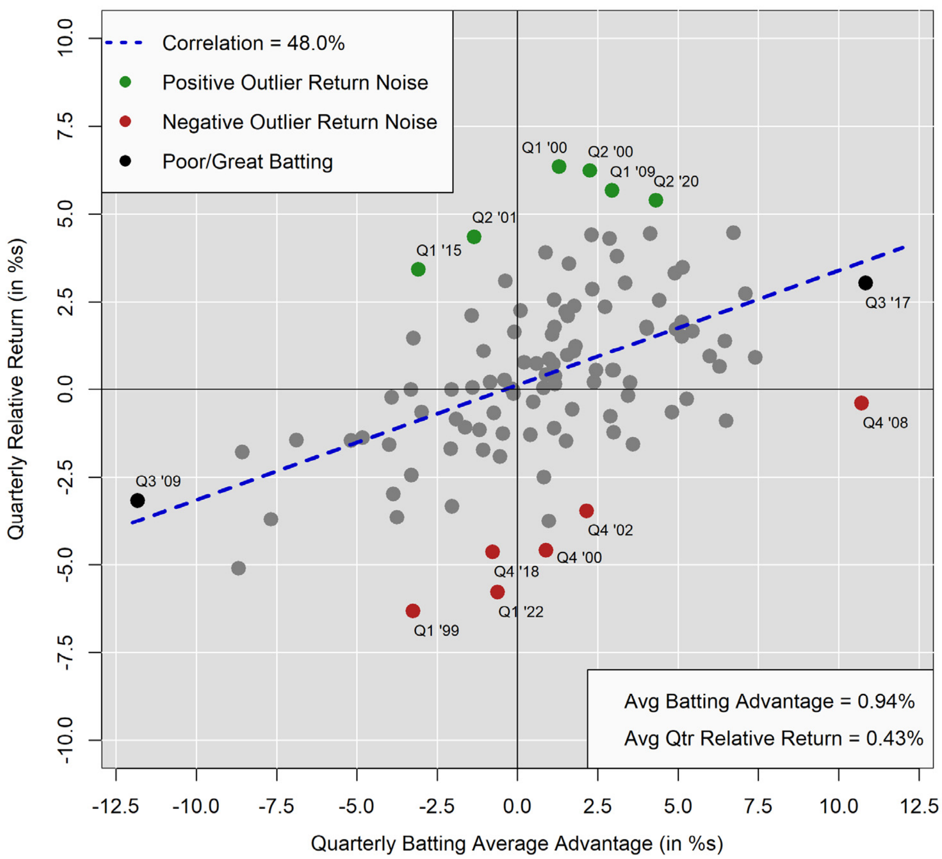
That the correlation of quarterly BAAs and relative returns is positive, but still well below a perfect correlation of 1.0, is what provides the extra stability to short-term performance evaluation. This benefit is similar to how in a single stock portfolio, adding a second holding can provide meaningful diversification benefits, even if the two assets are positively correlated with each other.

[6] The concepts of Omega and BAA are meant to be broadly applicable across managers and products. To avoid the distraction of specific strategy names, circumstances, and benchmarks, all the live data examples that follow have been anonymized.

Within Figure 1 there are a handful of quarters where realized returns greatly exceed the long-term trendline (points in green), or where returns are much worse than the relationship would predict (points in red). In both cases, keeping an Omega perspective helps avoid unproductive, reactionary thinking. When good return noise shows up in force, one can be thankful for the fortune but move forward without lifting their estimations for future returns. Perhaps even more importantly, when bad return noise strikes, BAA results can serve as a quick sanity check that the investment process is not necessarily broken and in need of repair.

Figure 1. Relative Returns vs. Batting Average Advantage

Performance results for "Portfolio X" relative to its benchmark index. Data covers October 1994 – March 2022.



Source: Compustat, FTSE Russell, and WEDGE Capital Management.

Insights at the Holdings Level

A natural worry of looking to binary performance measures is that they could miss out on a manager's expertise in locating big winners or avoiding big losers, either of which would have a powerful influence on portfolio returns. Fortunately, if holdings-level data is available to calculate BAAs, it is likely that these concerns can be addressed as well. Table 3 expands the BAA metric into various ranges of monthly stock returns for two separate, domestic equity products ("Portfolio Y" and "Portfolio Z," respectively) and their shared benchmark.

Table 3 is compiled from individual, one-month stock returns. Within each month, all holdings are standardized into the number of standard deviations[7] their stock return is from the overall benchmark's return. This standardized metric is known as a "Z-Score." Each month the Z-scores are grouped into eight buckets and tallied up. The buckets can be thought of as a baseball analogy for "singles," "doubles", "triples", and "home runs," in either direction. The monthly frequency tallies are then averaged across time, so Table 3 presents what the average month looks like for the portfolios and benchmark constituents. A true Normal Distribution is also shown for comparative purposes.

At least for these two products and their benchmark, most holdings tend to fall within ±1 standard deviation of the benchmark's return. Both products have shown positive BAAs on average, but that advantage is coming from two areas: having more positive "singles" in the portfolio, and avoiding big losers (see the far-right columns for a direct comparison against the benchmark's composition). In the "Home Run" outcomes, both products fall short of what occurs naturally in the benchmark. Thus, at least in this instance, the binary focus of BAA seems to be an appropriate measuring stick.

Table 3. Monthly Return Frequencies

Frequency of past one-month return outcomes for all holdings of "Portfolio Y," "Portfolio Z," and their shared benchmark index.
Data covers January 2009 – March 2022; frequencies shown in %s.

Return Z-Score		Average % of Holdings			Normal	Delta vs. Bench	
Min	Max	Port Y	Port Z	Bench	Dist	Port Y	Port Z
–	-3.00	0.25	0.14	0.61	0.13	-0.36	-0.47
-3.00	-2.00	0.99	0.52	1.63	2.14	-0.64	-1.11
-2.00	-1.00	6.03	4.29	7.66	13.59	-1.64	-3.37
-1.00	0.00	43.05	45.15	42.14	34.13	0.91	3.01
0.00	1.00	41.82	44.41	38.39	34.13	3.43	6.01
1.00	2.00	6.23	4.66	6.89	13.59	-0.67	-2.23
2.00	3.00	1.21	0.64	1.63	2.14	-0.43	-0.99
3.00	–	0.42	0.18	1.03	0.13	-0.61	-0.85
Batting Average		**49.68**	**49.90**	**47.95**	**50.00**	**1.73**	**1.94**

Source: Compustat, FTSE Russell, and WEGE Capital Management.
Note: Individual stock returns are standardized within each month into Z-scores based on standard deviations from the overall index return that month. Within each month frequencies are tallied for each of the Z-score outcome ranges, and then averaged across time for each Z-score outcome range.

[7] Standard deviation is first computed using the cross-section of index constituents, and then filled in for non-index holdings.

The data of Table 3 can be reorganized further into a comparison by magnitude of return outcome, where the Z-score ranges correspond to either positive or negative relative returns of that size. This reorganization offers a simpler comparison to a Normal Distribution. The middle columns of Table 4 reveal that the empirical distribution of stock returns has more singles and home runs than a Normal Distribution does, but fewer doubles and triples. More specifically, individual stock returns tend to be fat-tailed, or leptokurtic. This behavior is not a recent phenomenon: Eugene Fama wrote his PhD thesis on the topic in 1965, finding similar effects for the 30 Dow Jones Index members at that time.

The existence of these fat tails implies that for active managers, most available stock picking opportunities will offer only mild return outcomes, but there will be a few powerful outlier returns sprinkled across the universe from time to time. The low frequency of those big outliers means that even with a special skill to identify them more accurately, the return tailwind from outlier selection will only show up sporadically in an active portfolio, which creates an additional source of tracking error in returns. Omega avoids this extra layer of variance by scoring all outperforming stocks equally.

Table 4. Monthly Return Frequencies by Magnitude

Frequency of historical one-month return outcomes for all holdings of "Portfolio Y," "Portfolio Z," and their shared benchmark index.
Data covers January 2009 – March 2022; frequencies shown in %s.

Magnitude	| Return Z-Score |		Absolute Delta vs. Normal Dist			Net Delta vs. Bench	
of 1M Return	Min	Max	Port Y	Port Z	Bench	Port Y	Port Z
Singles	0.00	1.00	16.61	21.29	12.27	2.52	3.00
Doubles	1.00	2.00	-14.93	-18.23	-12.62	0.97	1.14
Triples	2.00	3.00	-2.08	-3.11	-1.01	0.22	0.12
Home Runs	3.00	–	0.40	0.05	1.37	-0.24	-0.38

Source: Compustat, FTSE Russell, and WEGE Capital Management.
Note: Individual stock returns are standardized within each month into Z-scores based on standard deviations from the overall index return that month. Within each month frequencies are tallied for each of the Z-score outcome ranges, and then averaged across time for each Z-score outcome range.

Omega Thinking

Once the basic framework of Batting Averages and Omega are understood, thinking in terms of probabilities and winners vs. losers can open the door to other interesting topics. For example, knowing that net of fee returns-based outperformance can come from a BAA as small as 1% – 2% helps frame various modeling and risk management decisions. At WEDGE Capital, quantitative models avoid relying on precise return forecasts and complex position weight schemes, instead favoring a more parsimonious approach of Buy/Hold/Sell classification rules and equal-weighted position sizes. Similarly, traditional research analysts are cautioned to remember that even a "high conviction idea" has a sizeable chance of not panning out. Therefore, the focus of a research analyst should be to maintain a healthy cadence of quality ideas, rather than scouring the landscape for big winners or surefire outcomes.

Concluding Remarks

While nothing will displace the relevance of studying past returns in active management, it is important to appreciate the amount of noise inherent in returns, and to seek out other data points that can expand the "scoreboard" of performance evaluation. One approach involves shrinking the fat tails of stock return outcomes into binary measures of whether a stock outperformed its benchmark or not. This framework of winner and loser probabilities can grow into broader, portfolio-level insights, such as the Batting Average Advantage metric, or more generally, the Omega dimension of performance.

As diverse measures of performance like Omega are added to the analytical toolkit, other interesting questions may percolate as well. A current focus area of research at WEDGE Capital asks how quantitative models and traditional research teams should be organized if Omega were the true, underlying heartbeat of generating skillful returns outperformance. See Bolshakov et al. (2021) for the latest academic publication on that topic.

Disclaimer

References

Arnott, R., Kalesnik, V., and Wu, L. (2017). The Folly of Hiring Winners and Firing Losers. *Journal of Portfolio Management.* 45 (1) 71-84.

Carhart, M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*. 52 (1), 57-82.

Cornell, B., Hsu, J., and Nanigian, D., (2017). Does Past Performance Matter in Investment Manager Selection? *Journal of Portfolio Management.* 43 (4), 33-43.

Cremers, M. and Pareek, A. (2016). Patient Capital Outperformance: The Investment Skill of High Active Share Managers Who Trade Infrequently. *Journal of Financial Economics,* 122 (2), 288-306.

Bolshakov, A., and Chincarini, L. (2020). Manager skill and portfolio size with respect to a benchmark. *European Financial Management,* 26 (1), 176-197.

Bolshakov, A., Chincarini, L., and Lewis, C. (2021). Enhanced indexing and selectivity theory. *European Financial Management,* 2021, 1-35.

Fama, E. (1965). The Behavior of Stock Market Prices. *Journal of Business,* 38, 34-105.

Fisher, R. (1934) *Statistical methods for research workers* (5th ed.). Edinburgh: Oliver and Boyd.

Fog, A. (2015). Biased urn theory. Working Paper, December 28. Retrieved from https://cran.r-project.org/web/packages/BiasedUrn/vignettes/UrnTheory.pdf.

French, K. (2022). Kenneth R. French – Data Library. Retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Harvey, C., and Liu, Y. (2018). Detecting Repeatable Performance. *Review of Financial Studies.* 31 (7), 2499-2552.

Kennedy-Shaffer, L. (2017). When the Alpha is the Omega: P-Values, "Substantial Evidence," and the 0.05 Standard at FDA. *Food and Drug Law Journal,* 72 (4), 595-635.

Krzyzewski, M. (2017, March 23). Basketball and Beyond with Coach K [Radio broadcast]. *SiriusXM.*

Reiss, M. (2017, March 23). Bill Belichick: 'Our guys played like champions in the last 20 minutes'. *ESPN.* https://www.espn.com/blog/boston/new-england-patriots/post/_/id/4802073/bill-belichick-our-guys-played-like-champions-in-the-last-20-minutes.

Wallenius, K. (1963). Biased Sampling: The Noncentral Hypergeometric Probability Distribution (Technical Report No. 70). Department of Statistics, Stanford University.

Appendix

Formal Description of Omega

Omega (the last letter in the Greek alphabet, resembling a lowercase 'w') is defined as the probability weight advantage a manager has for selecting outperforming stocks ("winners") versus underperforming stocks ("losers") (Bolshakov and Chincarini, 2020). This Omega expression of skill is constant, whereas underlying probabilities within a benchmark can shift over time and update as stocks are being picked. However, the simplest case of such probabilities can be tied back to Omega easily. In a benchmark with an equal number of winners and losers, and where a manager is only selecting a single stock, the following relationship holds:

$$\omega = \frac{Pr(X = win)}{1 - Pr(X = win)} \tag{3}$$

This Omega terminology stems from a branch of mathematics known as Urn Theory (see Fog, 2015), which in its simplest description is an exercise of sampling green and red M&Ms out of a candy jar. $\omega > 1$ implies that when the "manager" reaches their hand into the jar, they exhibit a slight force of attraction towards the green M&Ms, making them more likely to be chosen. The reasons for defining a probability weight advantage rather than just the probability itself are twofold: (1) the initial green vs. red populations may not be known, and might vary across jars; and (2) those populations change as sequential picks are made from the jar.

Regarding item (2), imagine that the manager has picked ten consecutive green M&Ms. The remaining pool of available candies at that point would tilt more towards red than usual – perhaps so much so that red becomes the more likely next pick. When the manager puts their hand in for another draw, an Omega value above 1 boosts the probabilities towards green. That boost may not be enough for green to overtake red as the probabilistic most likely outcome; Omega just tilts the odds in that direction.

Returning to investing, Batting Average metrics relay periodic probability information, whereas Omega describes a constant degree of skill. For a single manager or model, the balance of winners and losers in their "selection jar" should be relatively stable, as is the method for how stocks are chosen. In that case, the motivations for employing a probability weight advantage are not present, and studying BAAs alone can be perfectly sufficient for a robust analysis. For managers with meaningfully different holding periods and/or selection mechanisms though, the circumstances of their jars can be different, which is where Omega becomes a more appropriate measure. The precise mathematical details for mapping BAA to Omega are somewhat involved, but can be simplified down to the following two equations:

$$\widehat{\omega} = \frac{0.5 + b^*}{0.5 - b^*}; \tag{4}$$

$$b^* = W_{coef}\left(\overline{BAA}\sqrt{3\overline{hp}}\right)$$

Equation (4) defines the estimated Omega for a manager, $\widehat{\omega}$, as the ratio between a standardized batting average advantage, b^*, applied to winners (in the numerator) and to the avoidance of losers (in the denominator). In the absence of skill b^* is zero, and the Omega estimate reverts to its unskilled level of 1. b^* itself is defined as follows: a Wallenius adjustment coefficient, W_{coef}, multiplied by the average Batting Average Advantage sampled from the manager's history, multiplied by the square root of three times the manager's average investment holding period[8].

Taking these items in turn, the Wallenius adjustment coefficient corresponds to how sequentially stocks are chosen by the manager. Picking stocks one at a time to reach a fixed number of holdings offers a course-correction path through the underlying probabilities in the benchmark, just as picking ten red M&Ms in a row should tilt the next selection's probabilities more heavily towards green. The extra stability provided by this course-corrective nature is described by the Wallenius Noncentral Hypergeometric Distribution (Wallenius, 1963). Picking stocks in a single, bulk approach, where the number of holdings in the portfolio itself is a random outcome, does not provide that course-correction stability, and is described by the Fisher Noncentral Hypergeometric Distribution (Fisher, 1934). A full Fisher approach would imply $W_{coef} = 1$, while more Wallenius-like approaches call for values above 1 but less than the true ω could be. In practice most managers likely will operate somewhere in between those two extremes, and the $\widehat{\omega}$ estimate is not very sensitive to W_{coef} anyways, so sticking with a constant value of ~1.05 avoids ambiguity and aligns well with most BAA's in the 1% to 2% range.

The average Batting Average Advantage is simply the sample average across the time series of observed BAA's. Within the square root term, the average holding period adjusts for the fact that the same BAA is more powerful if it occurs over longer time intervals, as fewer outperformers tend to exist in a universe over longer measurement periods (e.g. if the horizon is long enough, only the generational compounders stand out as outperformers: these are the Apples, Microsofts, Wal-Marts, etc.). The average holding period and BAA measurement must be expressed in the same units: if the BAA time series is measured at monthly intervals, then the average holding period applied to the formula should in months as well. Finally, the coefficient of 3 shows up in the radical due to effects of the Central Limit Theorem on Urn Theory selection processes.

[8] Average holding period should correspond to initial purchases and final exits, exclusive of weight rebalancing trades.

Meaning of Expected Returns

A common area of confusion, where quantitative finance overlap poorly with regular vocabulary, involves the phrase "expected returns." Contrary to what it sounds like, expected returns does not mean a forecast for the next period, an average of recent results, or a belief of what one expects will happen sometime in the future. In mathematics, "expectation" means the probability-weighted average of all possible outcomes, or alternatively, what would happen on average across an infinite number of observations. This definition also serves as a mathematical operation, just like addition or subtraction, and is used as such throughout statistics and econometrics. "Taking expectations" is denoted by a capital E. Equation 5 shows the discrete event version, calculating the expected value across i possible outcomes with a probability assigned to each possibility. Equation 6 is the equivalent definition for a continuous function of probability densities, such as the Normal Distribution.

$$E(x) = \sum_{i} x_i \, Pr(x_i) \tag{5}$$

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx \tag{6}$$

While the true expected returns of investment processes cannot be known with certainty, they can be estimated and tested against prespecified levels of confidence. Estimate or not, the concept of expected returns also gives rise to a natural definition of risk. For any known distribution or data sample, variance can be defined using differences in expected values:

$$\begin{aligned} var(x) &= E[x - E(x)]^2 \qquad (7) \\ &= E[x^2 - 2xE(x) + E(x)^2] \\ &= E(x^2) - 2E(x)E(x) + E(x)^2 \\ &= E(x^2) - E(x)^2 \end{aligned}$$

For an easy example to conceptualize, consider rolling a single six-sided die. The expected value of a roll is just the average of the numbers 1 through 6, as they are all equally likely to occur. Using Equation 7 and the algebra beneath it, the variance of outcomes can be calculated easily as well.

$$E(roll) = \frac{1+2+3+4+5+6}{6} = 3.5 \tag{8}$$

$$var(roll) = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} - 3.5^2 = 2.92 \tag{9}$$

Interestingly, a die's expected value can never appear on an individual roll. It is impossible to roll 3.5, so every realized outcome has a nonzero occurrence of "noise," or an error term, around the expected value, just like the summary description of expected returns presented on page 2. For investments, even if one could know the true model or skill function that defines expected returns across securities, it would be perfectly reasonable to still have some amount of unpredictable, stochastic noise occur around that expectation every period. The purpose of Table 2 is to demonstrate how formidable random noise like that is in the field of investing.

For a Normal Distribution, which was used to generate Table 2, the expected value and variance happen to match whatever mean and standard deviation (squared) were chosen as parameters. In Table 2 these parameters were defined as mean outperformance of 2% per year, and a tracking error of 4% per year (or in units of variance, $16\%^2$).

To illustrate the magnitude of noise in investing, compare the Information Ratios between the theoretical Table 2 manager and the dice rolling exercise. In investing, an Information Ratio ("IR") divides the historical average outperformance by tracking error, which for a known, complete distribution of outcomes is identical to dividing expected value by standard deviation. In Table 2 the manager's IR is 0.5, whereas the dice-rolling IR is 2.05! In other words, even this theoretical manager with a respectable IR of 0.5 – which in practice cannot even be guaranteed like it is here – will face a relative volatility of outcomes more than four times as large as the easily imagined case of rolling dice over and over. Once that unrealistic skill guarantee is removed, and skill can only be estimated from a limited data sample, which might have fatter tails than a Normal Distribution, the volatility of active return outcomes would grow even wider.

This simple example is meant to be humbling, but not discouraging, for investment practitioners. Maintaining a healthy respect for the signal-vs-noise challenges in active management can help inform a range of process and implementation decisions. Within WEDGE that appreciation led to developing a complementary measuring stick of performance (Omega), and favoring parsimonious model structures for live investment products.