Validation and Evaluation of the Action Score

An Independent Exploration of Old School Value's Stock Grader "Action Score"

Written by George Wu, Jan 9. 2017

By George Wu / Jan 9, 2017

Table of Contents

By George Wu / Jan 9, 2017

Introduction

For those unfamiliar with the Action Score, please first read Jae's post (http://www.oldschoolvalue.com/blog/valuation-methods/action-score-quality-value-growth/) introducing the Action Score. Essentially, the Action Score is a summarized measure from 0-100 based on 9 quality, value, and growth metrics from each company's financial statements. Jae demonstrated in his backtest that by holding the top 20 ranked Action Scores per year from 1999-2015, one could have generated a compound annual growth rate (CAGR) of >20%!

To better understand the Action Score and to confirm his results, I explored the Action Scores to validate his results and to evaluate the Action Score as a trading system, in hopes of using the Action Score to supplement my own investing.

Validation of the Action Score

As an initial step, I first wanted to simply verify that the Action Scores were being calculated correctly. I started with the 9 quality, value, and growth criteria and applied the Action Score formula to the 17-year backtested dataset in Jae's introductory post. I was able to replicate all of the Action Scores, but I decided to make a few tweaks that seemed appropriate:

1. Price-to-Book needed to be recalculated for 2015 as it was incorrect
2. When ranking ties, I preferred using the average ranking instead of randomly ordered ranking (i.e., if two stocks have the same Piotroski F-Score of 5, instead of randomly ranking them as 1 and 2, both would be ranked as 1.5)

The next step was to calculate the performance results from holding the top 20 annually ranked Action Score stocks from the start to the end of the year. Even with the slight tweaks, my results were very similar to Jae's results for both the Full Universe (all stocks) and the Filtered Universe (no OTC, Financials, Miners, Utilities):

Backtested Performance – Average Annual Returns holding top 20 ranked Action Score stocks

YEAR	FULL UNIVERSE		FILTERED UNIVERSE	
	Validation	Original	Validation	Original
1999	21.8	25.5	-6.0	-7.0
2000	6.6	11.2	4.2	13.3
2001	43.1	43.0	42.3	39.4
2002	21.6	22.0	16.4	6.6
2003	234.6	230.1	84.2	83.5
2004	19.8	11.3	20.6	22.8
2005	20.2	14.0	20.3	23.3
2006	87.4	87.6	21.9	18.4
2007	7.0	17.2	-0.5	8.1
2008	-24.9	-28.1	-36.8	-31.3
2009	43.5	56.8	56.7	74.0
2010	10.4	8.3	25.0	21.3
2011	12.3	15.6	9.2	10.1
2012	16.1	15.4	14.9	18.4

2013	83.4	76.0	56.4	54.1
2014	16.9	18.0	18.2	28.0
2015	7.8	-0.35	6.1	18.5
CAGR%	**29.2**	**29.0**	**17.7**	**20.7**

The final CAGR for the Full Universe was 29.2% in my validation backtest compared to 29.0% in Jae's original backtest, and for the Filtered Universe was 17.7% in my validation backtest compared to 20.7% in Jae's original backtest. Some of the discrepancies were likely due to the changes I made in the validation backtest; this can be seen in, for example, in the 2015 average returns due to the different Price-to-Book values.

As Jae also mentioned, the Full Universe of stocks may not be viable to invest in as it contains a lot of OTC stocks with low liquidity making it difficult to buy stocks without substantial slippage and trading fees. Thus, the Filtered Universe is likely to be a closer representation of the actual investing performance that may be possible prior to trading costs.

Jae also showed that the Action Scores had superior relative performance compared to the S&P500 and Russell2000 indexes, which only had CAGRs of 4.8% and 7.5%, respectively. As an added control comparison, I calculated the performance results assuming we could equally invest in all stocks in each universe:

PERFORMANCE	**FULL UNIVERSE**		**FILTERED UNIVERSE**	
	Validation	All Stocks	Validation	All Stocks
MEAN$_{PER\ STOCK}$	36.9	31.2	20.8	15.1
MEDIAN$_{PER\ STOCK}$	13.1	1.1	8.9	2.2
%POSITIVE$_{PER\ STOCK}$	60.3	51.1	58.5	52.3
CAGR(%)	29.2	27.7	17.7	10.7

Surprisingly, I found that if we could invest equally in all of stocks in the Full Universe, we would have had a CAGR of 27.7%! First, it's not feasible to invest in that many stocks, but more importantly, it confirms that the Full Universe is not a reasonable approximation of the set of truly investable stocks. When we turn to the Filtered Universe of stocks, we find that the observed CAGR of 10.7%, when investing in all stocks, is more in line with the S&P500 and Russell2000 CAGRs of 4.8% and 7.5%, respectively. In this Filtered Universe, Action Scores still outperform all stocks with a CAGR of 17.7%.

I added a few other measures of interest including the mean annual return per stock, the median annual return per stock, and the % of positive annual stock returns. For each of these measures in both universes of stocks, the Action Scores also outperformed compared to all stocks in the universe.

The mean return per stock was also substantially higher than the median return per stock. This is likely due to some stocks that have outlier-type annual returns (>>500%) that drag up the mean, while the median is simply the return of the 'middle' stock when stock returns are ranked. This allows the median to be a more robust estimate of the expected return per stock as it is unaffected by outliers. It is reassuring to find that both the mean and median of the annual return per stock is reasonably high, since it then becomes much less likely that the CAGR is driven by spectacular returns from only a few stocks. This can be most easily seen by comparing the mean and median returns per stock from the validation backtest using top 20 Action Score stocks annually to all stocks in the Full Universe.

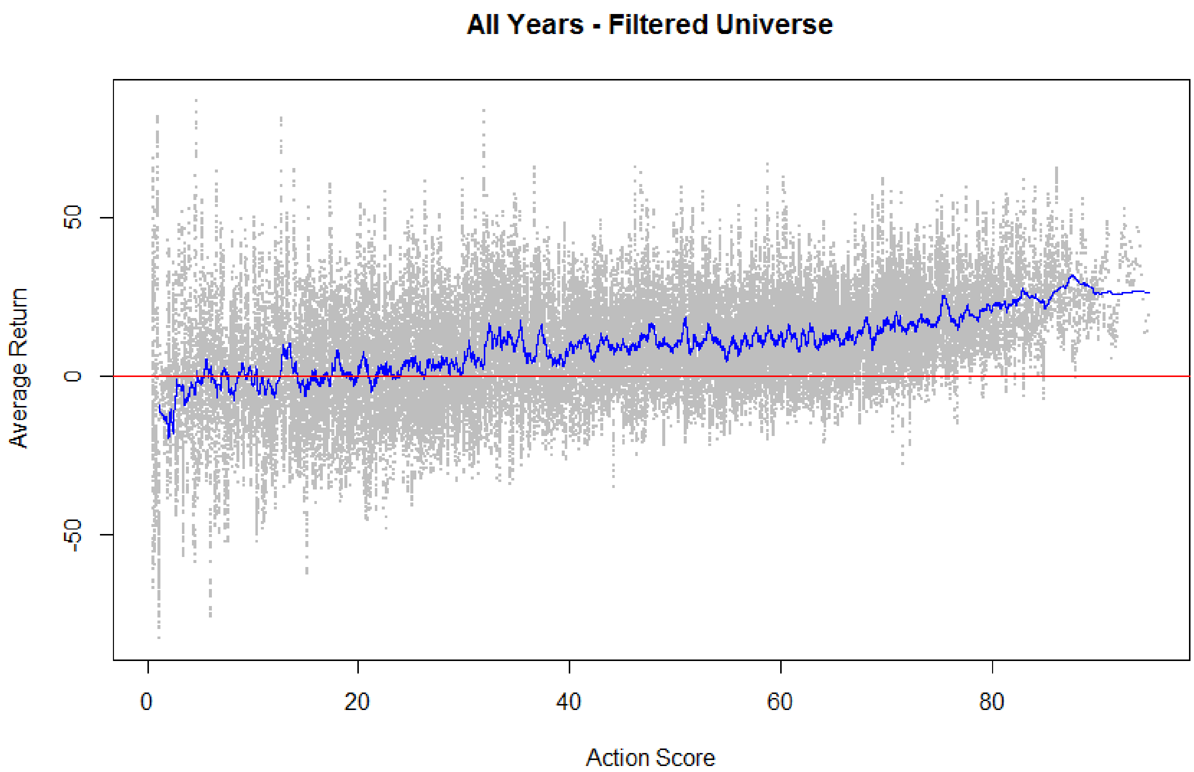
The results from the validation process confirm Jae's findings from his original backtests. Each Action Score calculation was verified, and the performance results when holding the top 20 ranked Action Scores per year were very similar with the results from my backtests. Additional measures also support that hypothesis that the Action Scores are indeed selecting a subset of stocks more likely to have higher annual returns. The Filtered Universe of stocks was found to be a better universe of stocks to use as a representation of the investable universe, thus throughout the rest of this report, for simplicity, only results using the Filtered Universe will be presented.

Evaluation of the Action Score

Having validated the Action Score and backtested their performance, I wanted to better understand the holistic performance of the Action Scores, rather than focusing only on the subset of annually top 20 ranked Action Score stocks. If the Action Scores represented a fundamental association between superior financial statements and higher annual stock returns, then we should expect to see that as Action Scores increased through the full range from 0-100, the annual stock returns should increase on average as well.

One way to visualize this is to rank all stocks in the Filtered Universe, including all years, by their Action Score, and then plot the average annual returns using a sliding window of 20 stocks from the lowest 20 Action Score stocks to the highest 20 Action Score stocks (note: 20 stocks is chosen simply for consistency and represents a reasonable number of stocks for an individual to invest in a year; results are similar with larger windows of stocks). To make the plots more manageable, returns were restricted to a maximum of 200% to deal with extreme outlier (>1000% return) stocks.

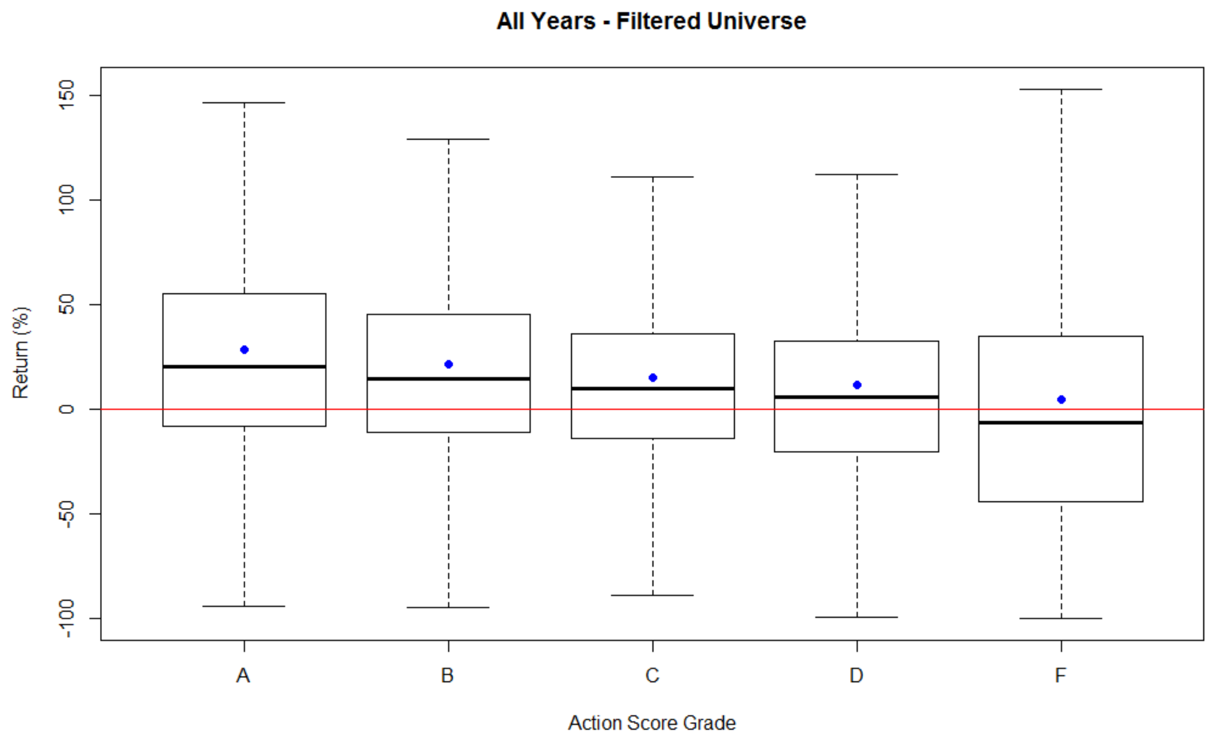
By George Wu / Jan 9, 2017



Each gray point is the average annual return for a single selection of 20 stocks going from the lowest 20 Action Score stocks to the highest 20 Action Score stocks. As the variability in the returns were quite large, I applied a moving average (500-unit window) across the Action Scores to smooth out the returns in order to capture the general trend (shown by the blue line). We can then clearly see that the average annual returns gradually increase as Action Scores increase, with the lowest amount of variability (spread of gray points) and highest overall returns in the upper ranges (>80) of the Action Score. In fact, there are very few gray points below the red line (average annual return of 0%) for Action Scores above 80, while in contrast, there are much larger proportions of gray points below the red line for Action Scores below 20.

Another way to visualize this is to group Action Scores across all years by grade from Grades A (>85), B (75-85), C (65-75), D (50-65), and F (<50), and then use a boxplot to summarize the distribution of returns for each grade. This is similar to a prior post by Jae comparing returns across grades with the intent to show that Action Scores with high grades do better than Action Scores with lower grades (http://www.oldschoolvalue.com/blog/valuation-methods/osv-rating-system-analysis/).

By George Wu / Jan 9, 2017



Returns (%) By Action Score Grade (Filtered Universe – All Years)

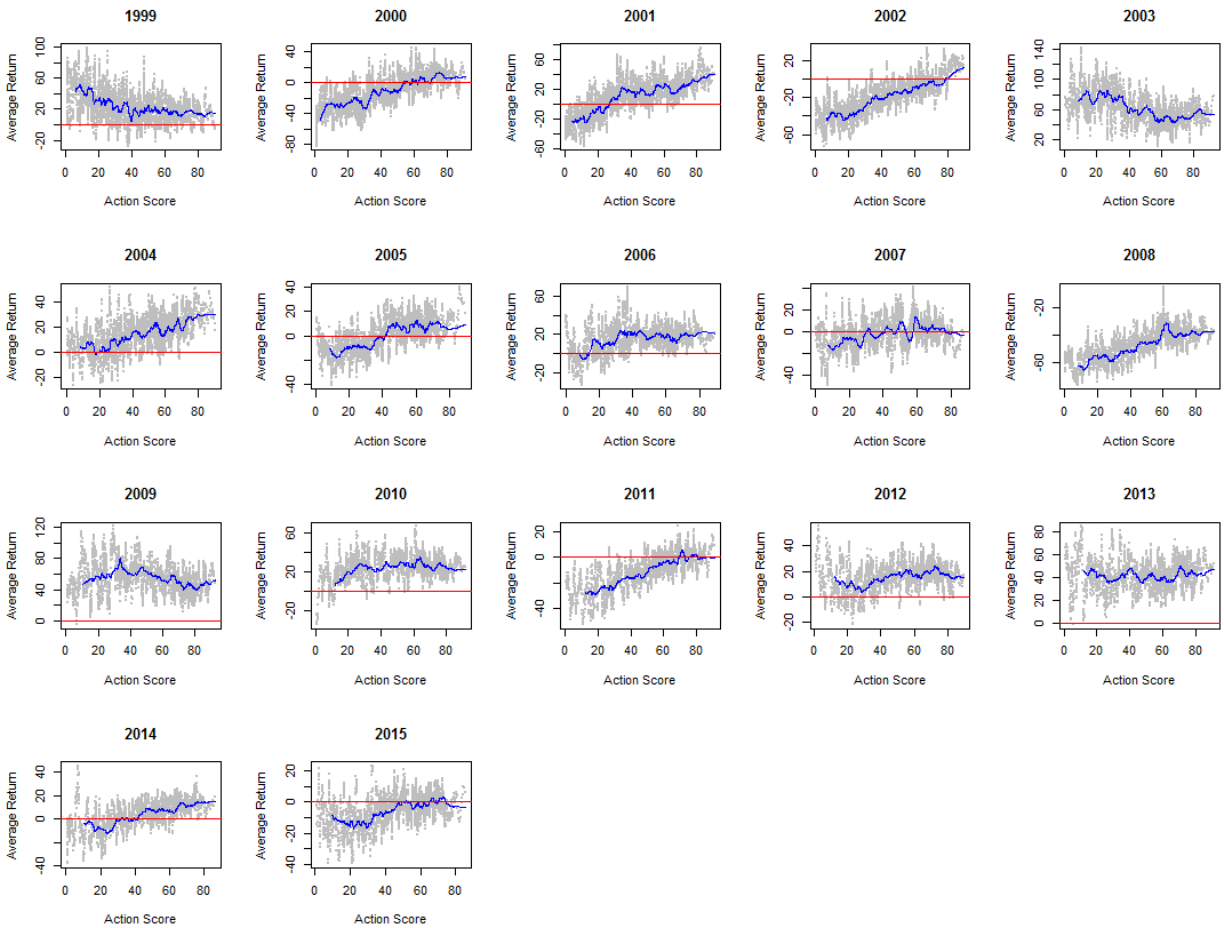
Grade	A	B	C	D	F
N	979	4417	9750	15175	34503
Mean	31.4	24.4	16.7	14.5	13.3
Median	20.6	14.7	9.9	5.7	-6.0
%Positive	68.1	65.4	61.7	56.8	45.5

The main features of the boxplot above is the solid black lines depict the median return of stocks in each grade, and the blue dots represent the mean return of stocks in each grade. We can see in the plot and the table that for both the mean and the median, grade A stocks have higher annual returns than grade B stocks, and similarly for grade B > grade C, grade C > grade D, and grade D > grade F stocks. This confirms the trend that we previously observed where more generally, the higher the Action Score, the higher the overall annual return (in the prior case for groups of 20 stocks ranked by Action Scores, and here for individual stocks). The top of the boxes and the upper whiskers (the dashed and connecting dotted line) show that there are more grade A stocks with high returns (eg, >50%) compared to the other grades. This suggests that a larger proportion of grade A stocks are more likely to have outsized positive returns compared to the lower grade stocks.

The above analyses help demonstrate that in general, higher Action Scores tend to be associated with higher annual returns across the full range of Action Scores and across Action Score grades.

By George Wu / Jan 9, 2017

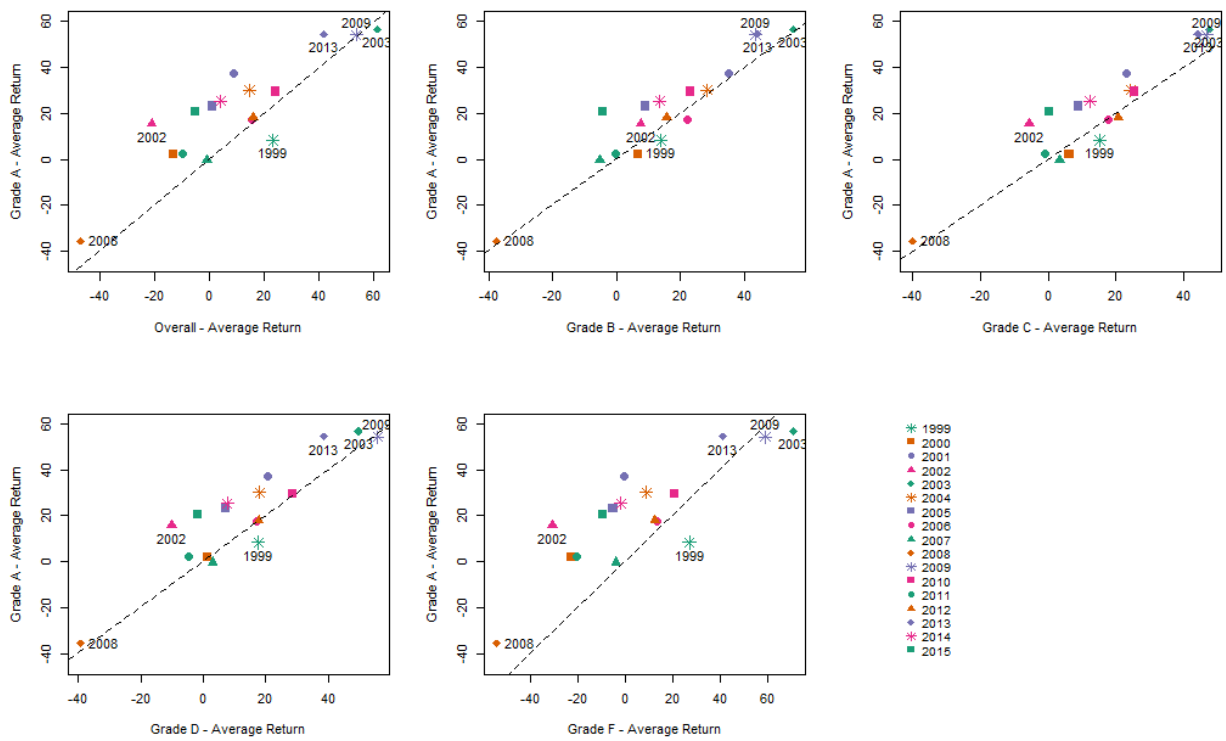
A natural follow-up question is whether this finding remains true, not only when all years are combined together, but also when breaking down the data by each individual year. This is important because in practice, we would use Action Scores to rank stocks within a given year and then select stocks for each year based on those rankings, rather than trying to select for the highest ranked Action Score stocks across all years at once.



To answer this question, we can repeat the same analyses as before, but broken down by year. I personally like the average annual returns by Action Score plots the best, so I will use it to illustrate the findings. Focusing on the blue general trend lines, we see confirmation of the expected relationship between Action Scores and annual average returns for most of the years, like for 2001, 2002, or 2011. We also see years in which the relationship seems to be almost non-existent or reversed like for 1999, 2003, or 2013. If we look closer, we find that the years in which the trend is weak or reversed correspond to years in which the overall stock return is very high. For example, in 2003, the average stock performance for any selection of 20 stocks is almost all above an average return of 20%. In contrast, for 2008 or 2002, where overall stocks performed much worse, the Action Scores do a better job of ranking average returns. Ideally we would like to see that higher Action Scores are associated with higher average annual returns consistently across all years, but that does not appear to be true. The silver lining in this finding is that if the Action Scores are not ranking stocks properly, chances are any selection of stocks for that year will still have performed well.

By George Wu / Jan 9, 2017

Another minor point to make is that the 0% return line in red jumps all over the place, and the y-axis needs to change scales a lot to properly plot the average returns for each year. This variability suggests that although higher Action Scores in general seem to be linked to higher average returns within a given year, the actual performance of higher ranked Action Scores is still primarily driven by the overall stock performance for that year. This can be hard to conceptualize, so I created another set of plots that focus on the relative performance of Grade A stocks versus other Grade stocks and to all stocks overall within each year.



In each of the scatterplots above, each individual point is the average return of grade A stocks in a given year compared to the average return of all stocks (top left), grade B stocks (top middle), grade C stocks (top right), grade D stocks (bottom left), and grade F stocks (bottom middle) in that year. The most important line in the plots is the dashed diagonal line, which indicates when the average returns for the two grade categories would be equal. As an example, using the top left plot which compares grade A stocks to all stocks overall, we find that almost all points are above the dashed diagonal line, which means that in each year, grade A stocks tend to have higher average returns than all stocks overall. This finding is also observed in all of the other plots, indicating that for each year, average returns for grade A stocks also tend to be higher compared to lower grade stocks.

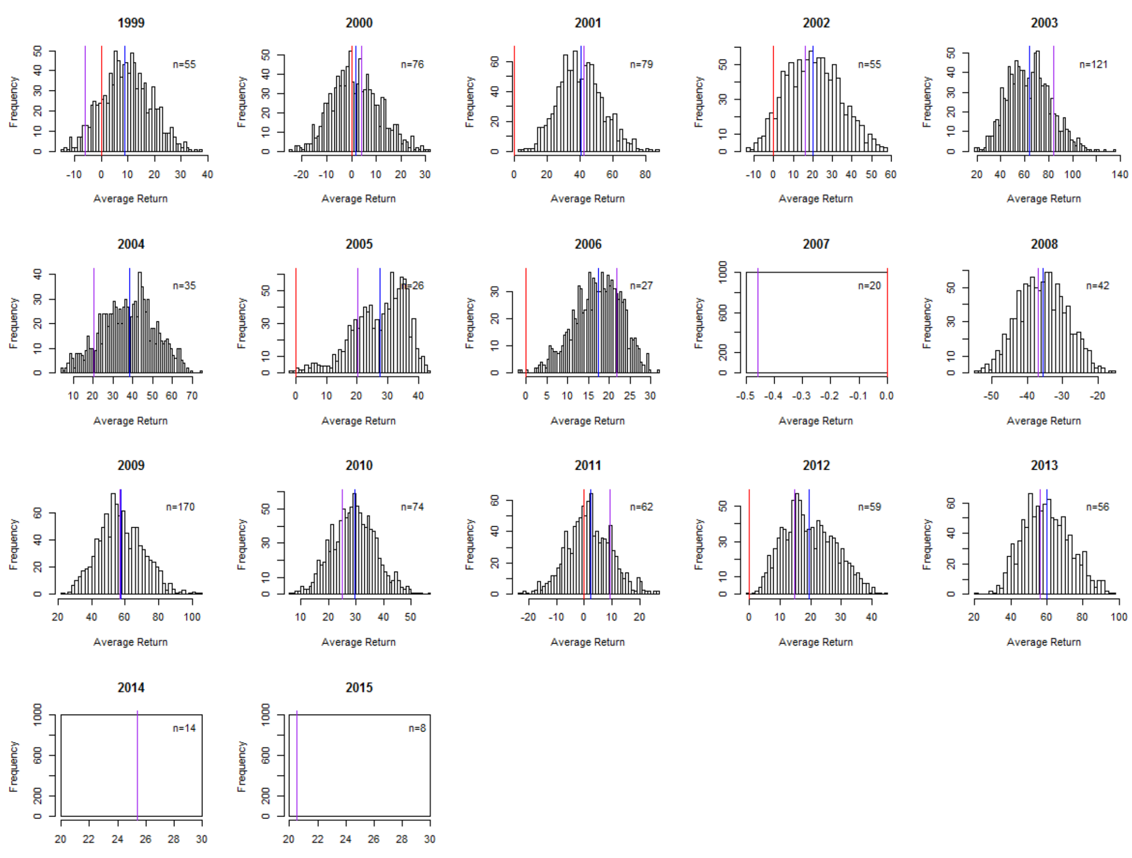
A few specific years are highlighted on the plots, including the most recent major recession year of 2008, where although grade A stocks did do better than lower grade stocks, the return (<-30%) was extremely poor. 2002 represents the more common result, where grade A stocks outperform all stocks overall and lower grade stocks in that year. For 1999, we see that grade A stocks performed worse than lower grade stocks, and similarly, in 2009 and 2003, grade A stocks underperformed compared to lower grade stocks, but notice that in the latter two years the returns on average for all stocks was >40%. This confirms our previous finding that higher Action Scores do not always equate to higher returns for a given year, but when that does occur, the overall average return for all stocks for that year tends to be high.

Another point worth re-emphasizing is that high (eg, grade A) Action Scores will not protect your investments from a market crash. Expected investment performance is still primarily driven by the overall equity market performance for that year, as annual returns are more similar across years than across grade categories.

Simulations

At this point, I'm reasonably convinced that higher Action Score stocks, in particular Grade A stocks, are likely to provide higher relative performance in the backtested data. The question I wanted to ask is if I employ this strategy by selecting 20 stocks myself from the available Grade A stocks in a year, how would my portfolio performance vary?

To answer this question, I first simulated the different average returns we might observe by randomly selecting 20 Grade A stocks in each year 1000 times and then calculating the average return. For years in which there were 20 or less Grade A stocks, there would only be 1 possible average return calculated using all of the available Grade A stocks (this occurred in 2007, 2014, and 2015). As for the prior plots, we restricted returns to a maximum of 200% to more easily deal with outlier returns obscuring the plots.

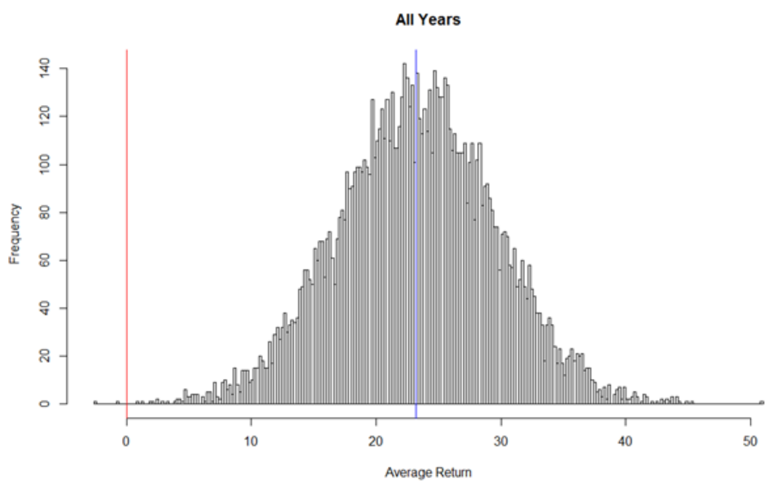


By George Wu / Jan 9, 2017

To visualize the distribution of returns, I used histograms, which show the number of times out of the 1000 tries per year that a random 20-stock holding fell into each bin of average return ranges, with the red line marking an average return of 0%, the blue line marking the average return of all grade A stocks, and the purple line marking the average return of the top 20 ranked grade A stocks per year.

The average returns were quite variable, such that if we had selected very poorly, it is possible to have negative returns in many of the years. On the flip side, the overall grade A stock average return fell in the middle of the histogram, suggesting that there are also many ways in which we could have selected 20 grade A stocks that delivered higher average returns. In particular, within each year, the top 20 ranked grade A stocks do not seem to show any association with the highest average returns possible, suggesting that there is room for us to add value using our own investing acumen to choose the best 20 grade A stocks to invest in.

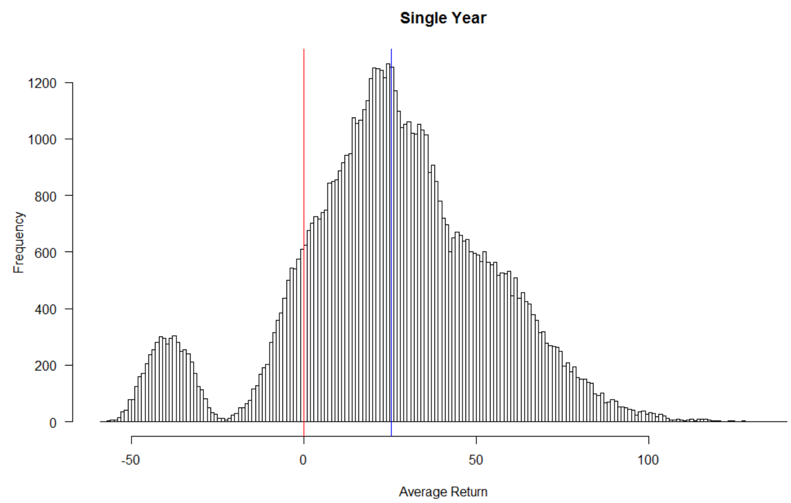
My next step, to more directly answer the question, was to randomly choose 20 Grade A stocks for each year, calculate the average return in that year, repeat for each of the 17 years, and then calculate the average of those average yearly returns across the full 17-year period. This is analogous to simulating the returns from the repeated construction of different 17-year portfolios of annually 'self-ranked' top 20 Grade A stocks. Repeating this 50000 times, I observed the following distribution of returns:



Average Return Per Year	Probability (50000 simulations)
< 0%	0.0 %
>5%	99.8 %
>10%	98.2 %
>20%	69.6 %
>30%	14.2 %
>40%	0.4 %

The blue line represents an average return of 23.6% across 17-years when randomly holding 20 grade A stocks per year. The simulation indicates that there is a very low chance of obtaining less than 10% average yearly returns (similar to a CAGR near 10%) when randomly selecting 20 stocks annually among Grade A stocks for 17 years. More than half the time, we would expect to have >20% average yearly returns, and for some lucky portfolios (>14% chance) of >30% average yearly returns!

It is important to note that the above simulation would apply, only if we observed the exact same distribution of returns for grade A stocks and for each of the 17 years we invested in. A more practical simulation may be to randomly select a year, and then randomly select 20 stocks to hold, to see what our expected performance would be for any given year. This allows us to see the distribution of single year returns, as we probably won't know whether stocks will over perform (like 2003) or under-perform (like 2008) in the coming year.



Average Return Per Year	Probability (5000 simulations)
<-10%	7.4 %
< -5%	9.3 %
< 0%	19.0 %
>5%	76.3 %
>10%	70.6 %
>20%	56.2 %
>30%	36.0 %
>40%	24.6 %
>100%	0.2 %

Making this adjustment, we see that the average return for a single year is 23.2% (blue line). We also find there is a much higher chance of having a negative year (19%), although more than half the time we would expect to have >20% returns.

We also see a separated group of negative returns, which turn out to be mostly a result of randomly selecting 20 stocks from 2008. This reinforces that idea that investment performance using grade A stocks for a given year is primarily driven by the overall stock market performance for that year. Hence, investment performance in the future using Action Scores are unlikely to replicate the results seen in the 17-year back test, since we cannot assume that performance of equities will be the same in the coming 17 years. If we have more years like 2008 in the future, then we will almost certainly have much lower returns.

The key difference between this simulation and the prior simulation is the current simulation showed the distribution of returns for a <u>single</u> year, while the prior simulation showed the distribution of average returns over <u>17</u> years. This means that there is a much higher chance of poor returns, since the returns aren't being averaged out by the other 16 years, which will all have higher returns than 2008. An alternative way to state this is that the average returns for one year will be much more varied than the average return across multiple years, reflecting the general principle that as investors we are more likely to see a better estimate of our 'true' investment performance if we have invested for many years.

Test Portfolio

Finally, I wanted to walk through the investing experience of using Action Scores by imagining the daily emotions that may transpire as I monitored my portfolio daily over the full 1-year holding period. The importance of this last experiment can't be understated, as even if we did have a system to generate an investing edge, we also have to be able to hold our investments through the swings in prices.

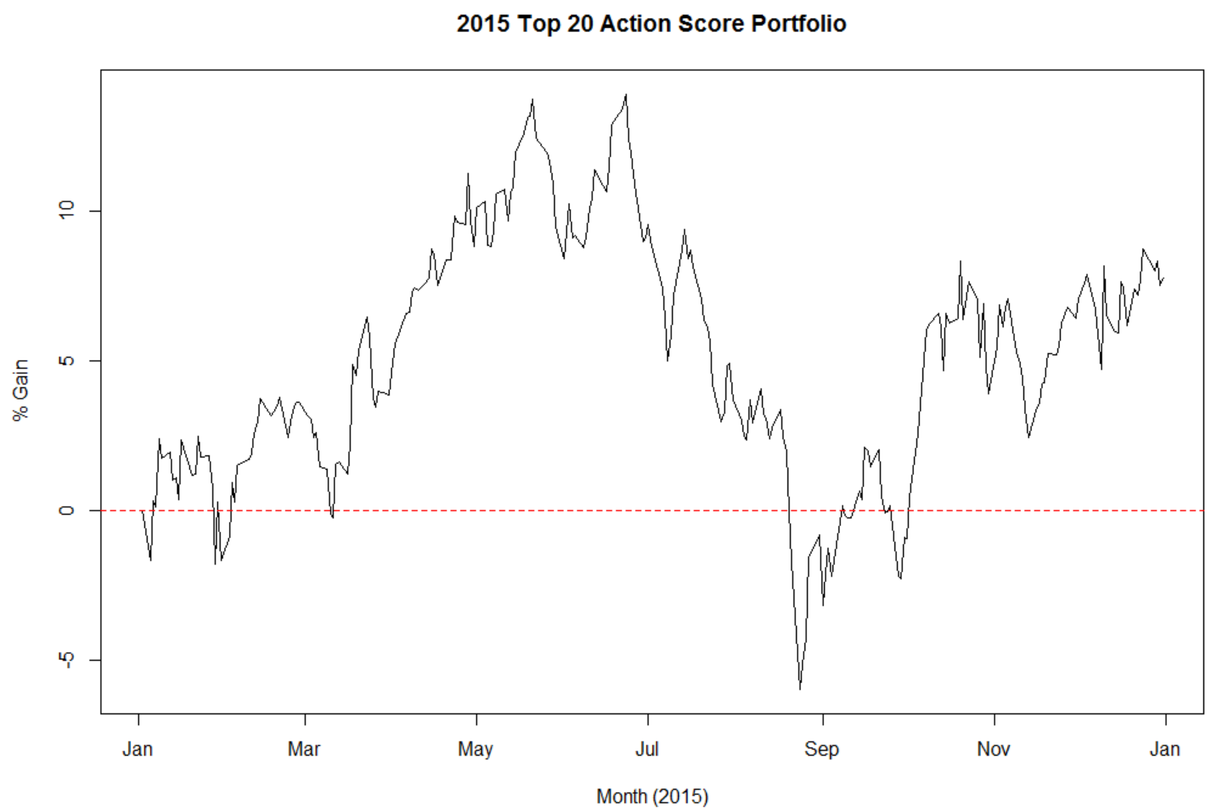
I decided to use 2015 as the experimental year. I first used the Action Scores to identify the top 20 ranked Action Scores for 2015 to invest in, and then recorded the results assuming I bought the stocks on 2015-01-01 and then sold them on 2016-01-01. The stock holdings in my portfolio and eventual returns are shown below:

Ticker	Company Name	Year	Start	End	Return	Action
LRN	K12 Inc	2015	11.7	8.8	-24.8%	91.3
BCOR	Blucora Inc	2015	13.51	9.8	-27.5%	88.7
FONR	Fonar Corp	2015	10.26	17.26	68.2%	88.3
TTWO	Take-Two Interactive Software Inc	2015	28.1	34.84	24.0%	87.6
NTT	Nippon Telegraph & Telephone Corp Ntt	2015	25.63	39.74	55.1%	87.2
HLF	Herbalife Ltd	2015	37.58	53.62	42.7%	86.7
ONE	Higher One Holdings Inc	2015	4.22	3.24	-23.2%	85.3
KZ	KongZhong Corp	2015	5.56	7.5	34.9%	85.2
MOC	Command Security Corp	2015	1.75	2.31	32.0%	84.8
UWN	Nevada Gold & Casinos Inc.	2015	1.22	2.25	84.4%	84.7
OUTR	Outerwall Inc	2015	71.52	36.54	-48.9%	84.5
STS	Supreme Industries Inc.	2015	139.25	128	-8.1%	83.9
PPC	Pilgrim's Pride Corp	2015	31.66	22.09	-30.2%	83.7
MBT	Mobile TeleSystems PJSC	2015	7.17	6.18	-13.8%	83.6
THTI	THT Heat Transfer Technology Inc	2015	1.21	0.48	-60.3%	83.5
OVTI	OmniVision Technologies Inc	2015	26.02	29.02	11.5%	83.4
SANM	Sanmina Corp	2015	23.37	20.58	-11.9%	83.3
ACCO	ACCO Brands Corp	2015	8.79	7.13	-18.9%	83.2
STRZA	Starz	2015	29.44	33.5	13.8%	82.8
ARRS	ARRIS International plc	2015	30.25	30.57	1.1%	82.4
				Average	**5.0%**	

2015 would have been a positive year, with an average annual return per stock of 5.0%, and exactly half of the top 20 stocks would have had positive returns. In comparison, the annual return of SPY was 1.89% and for Russell2000 was -3.42%. Note, the average return for 2015 is slightly different (5.0% vs 6.1%) than reported in the validation analysis described previously, as I assumed different entry and exit dates due to using other data sources to extract the daily prices.

By George Wu / Jan 9, 2017

For those who invested in 2015, you will know that 2015 definitely had its shares of ups and downs as the stock market experienced a rapid correction triggered by the Flash Crash in August. Thus, we can expect the top 20 Action Score portfolio in 2015 also experienced similar high volatility. This was indeed observed as seen in daily portfolio returns:



At the end of the year, the portfolio resulted in a **7.8%** gain not factoring in commissions, slippage, or other trading fees. But in order to achieve that gain, we had to sit through a ~20% drop in our portfolio from the highs in mid-June to the lows in late-August! As an investor, I would need to have nerves of steel to resist bailing ship on the Action Score system in late-August, as fears of a serious market crash would haunt my dreams. To further illustrate this stress, I calculated the max total gains and losses for each individual stock in our 2015 portfolio sorted by annual return:

By George Wu / Jan 9, 2017

Ticker	Action	Max Gain (%)	Max Loss (%)	Return
UWN	84.7	104.9	-17.2	84.4%
FONR	88.3	80.1	-11.3	68.2%
NTT	87.2	57.2	-2.2	55.1%
HLF	86.7	64.8	-26.6	42.7%
KZ	85.2	54.3	-12.2	34.9%
MOC	84.8	102.9	-23.4	32.0%
TTWO	87.6	31.7	-17.1	24.0%
STRZA	82.8	58.3	-7.2	13.8%
OVTI	83.4	13.3	-23.1	11.5%
ARRS	82.4	24	-19.1	1.1%
STS	83.9	6.3	-13.5	-8.1%
SANM	83.3	9.7	-24.9	-11.9%
MBT	83.6	78.4	-15.5	-13.8%
ACCO	83.2	4.7	-22.6	-18.9%
ONE	85.3	5.2	-56.2	-23.2%
LRN	91.3	51.4	-25.2	-24.8%
BCOR	88.7	23.3	-34.5	-27.5%
PPC	83.7	17.4	-45.1	-30.2%
OUTR	84.5	19.2	-49.0	-48.9%
THTI	83.5	2.5	-62.0	-60.3%

Would I be able to sit through >20% losses to obtain >30% eventual returns for stocks like HLF and MOC? Or would I excessively worry about stocks like PPC, THTI, or OUTR, which had >45% max losses? The lesson I think to be learned is that even if we believe Action Scores offer us an edge in predicting stocks with higher annual returns, we as investors need the discipline to hunker down through the storm until it ends. If we aren't able to do so, then the variability inherent in the annual buy and hold method of the Action Score system would for sure lead us to losses or, at best, substantial underperformance, since we'd always be selling at the lows.

Final Thoughts

After thoroughly exploring Action Scores, I feel confident in stating that Action Scores have the potential to identify stocks that will relatively outperform in terms of annual returns for a given year. From the 17-year back test, the key takeaways from my analysis were:

- Action Scores in the 17-year backtest confirmed Jae's posted returns of >20% CAGR
- In general, higher Action Scores were associated with higher annual stock returns
- This was not true, specifically for years in which the stock market overall did very well
- Grade A Action Score stocks tended to outperform compared to lower grade Action Score stocks
- The primary driver of Action Score stock performance was the overall stock market performance for a given year
- Simulations showed that there was a wide range of possible returns when randomly holding 20 grade A Action Score stocks per year with a mean average return per year of >20% for both single year and 17-year portfolios

- High variability in daily portfolio returns may be observed when employing the system in practice, especially for daily individual stock returns

I think the analyses convinced me that Action Scores are definitely worth considering when investing, especially focusing on the subset of Grade A Action Score stocks in each year. The cherry on the top of the cake would be if we could apriori identify the years in which stocks overall will perform poorly, which will then allow us to avoid massive down years like 2008, but that would be a discussion for another time.

One final very important point is that the conclusions from this report are only specific to the 17-year backtest data. The simulations performed assumed the same distribution of annual returns as 1999-2015. The Action Scores were developed using the same dataset as which was used to evaluate their performance, which means that the reported results **still need to be verified in an independent dataset**. For the development of any quantitative system, this last step is absolutely crucial because if we search hard enough across a wide array of possible features, we can always find patterns that result in high returns that back-test well. To truly validate the Action Scores and their performance, in hopes of extrapolating the findings into the future, it is necessary to analyze how the Action Scores would perform in independent data. Fortunately, I'm sure Jae will update on the Action Score results for 2016 and future years, and by observing the future returns, we may be able confirm the findings from this 17-year back test. But even if the results are confirmed positively, there is always a risk that any edge provided by Action Scores could disappear, or that the overall stock market experiences a major bear market, which means investing in equities are likely to generate poor returns even when using the Action Scores.

I wish everyone happy investing and hope your portfolio generates fantastic returns in 2017!

About the Author

My name is George Wu. I'm a casual investor with a quantitative background in statistics. My investing journey began in 2013 with classic investment books like the Market Wizard series and an initial dive into financial research, where I uncovered Novy Marx's 'The Quality Dimension of Value Investing' paper. Attracted to the idea of quality, value, and momentum ever since, I found myself developing as a fundamental investor. I stumbled on Jae's blog awhile back and felt he had an interesting system worth exploring, so I subscribed and contacted Jae trying to learn more about the Action Score system. Jae was kind enough to eventually let me play around with the data myself, and from those efforts came this report.