To Equal Weight Or Not To Equal Weight: That Is The Question.

Peter A. Forsyth*

November 27, 2018

Warning: this blog is a bit technical. However, without some pain, there is no gain.

The Question

There is some evidence that an equal weighted stock index may be superior to the common capitalization weighted indexes [1]. A capitalization weighted index (cap-weighted for short) simply weights each stock in the index in proportion to the total market value of its outstanding shares. Many of the popular ETFs track market cap indexes, such as the $S\&P$ 500 or the TSX 60.

On the other hand, an equal weighted index invests an equal amount of cash in all the stocks in the index, and then rebalances the portfolio frequently. This is an extreme form of diversification. There is some evidence that this strategy captures many of the *smart beta* effects that are now being hawked to investors. For example, an equal weight strategy will clearly invest more in small cap stocks compared to a cap-weighted index. Small capitalization stocks have been suggested as one of the *factors* in smart beta strategies. I think of an equal weighted index as a *dumb smart beta* strategy.

In my previous blog [2] I compared some investment strategies using a conventional measure of performance, the Sharpe Ratio. However, this is a fairly simplistic measure of risk and reward. The Sharpe Ratio measures risk in terms of the average volatility of the investment portfolio. One of the flaws of using volatility as a measure of risk is that it penalizes gains as well as losses with respect to the average return. I am reminded of a comment from a banker sitting next to me on a flight from New York: *"Actually, we like volatility when the price of the stock goes up."*

The Bottom Line

To summarize, the purpose of this blog is to examine the following question

> *"Does an equal weight index ETF outperform a comparable market cap weighted ETF by any measures of value to investors?"*

Bootstrap Resampling

So, in order to evaluate an investment idea, we need to gather some more useful statistics. A popular technique is based on a *data driven* approach. Here, we make no theoretical assumptions

*David R. Cheriton School of Computer Science, University of Waterloo, Waterloo ON, Canada N2L 3G1, paforsyt@uwaterloo.ca, +1 519 888 4567 ext. 34415.

[1] See "Why Does an Equal-Weighted Portfolio Outperform Value and Price-Weighted Portfolios?" authors: Y. Plyakha, R. Uppal, G. Vilkov, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2724535

[2] "Bonds in a balanced portfolio: long term or short term?" https://www.pwlcapital.com/bonds-balanced-portfolio-long-term-or-short-term/

about how markets behave, we simply use the historical data directly. Suppose we are interested in examining a long-term strategy, which would be typical of a person saving for retirement. In this case, for a 35-year old investor, planning to retire at 65, the investment time frame would be 30 years.

The most reliable historical data goes back to about 1926. One popular method for evaluating an investment strategy is based on starting the strategy in 1926, and then seeing how well the portfolio did 30 years later, i.e. in 1956. Then, we do the same thing starting in 1927 and ending in 1957, and so on. The problem with this rolling year approach, is that there is a large overlap between each of the rolling year samples, so we can't use this to generate reliable statistics.

At the other extreme, we can use a *bootstrap resampling* approach. A thirty year investment scenario consists of 360 consecutive one month returns. A single scenario is constructed as follows. We select a month at random from the historical data, and use this as our first month's return. Then, we select another month at random [3], which is the second month's return in our thirty year scenario. We keep doing this until we have a set of 360 returns (one thirty year path). We then repeat this procedure many times, to produce many 30 year return paths.

However, this bootstrapping approach does not take into account possible *serial correlation* in the returns. This is just a another way of saying that next month's returns may be affected by the returns of the past few months or years.

To take this into account, we select an initial month at random, but use b consecutive monthly returns (starting at the initial random month). We repeat this $(360/b)$ times to generate a single 30 year path. We call b the blocksize.

But we are not done yet. It turns out that a better approach is to not use a fixed blocksize, but to specify an average blocksize b, and randomly vary the blocksize within each thirty year path. This is called the *stationary block bootstrap* method. [4]

Now, armed with our bootstrap resampling technology, we can proceed with asking some interesting questions. [5]

The Data

Our historical data covers the time frame 1926:1 - 2015:12. Most of the data was obtained from the Center for Research in Security Prices (CRSP), on a monthly basis.[6] For the stock market data, we use the CRSP value (capitalization) weighted total return index. This index includes all distributions for all domestic stocks trading on major U.S. exchanges. We also use the CRSP equal weighted index, which is rebalanced monthly. The bond index is the US 10-year Treasury index.[7] All of these various indexes are in nominal terms, so we adjust them for inflation by using the U.S. CPI index, also supplied by CRSP. We use real indexes since long term investors should be focused on real (not nominal) wealth goals. All results will be reported in constant US dollars, and we assume no fees, taxes or transaction costs.

[3] There is a small chance we select the same month. We are sampling with replacement.

[4] "The stationary bootstrap." Journal of the American Statistical Association, authors: Politis, Romano; vol. 89 (1994) 1303-1313.

[5] Bootstrap resampling can be used to provide training data for machine learning methods, see "A data driven neural network approach to optimal asset allocation for target based defined contribution pension plans," Y. Li and P.A. Forsyth, `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3192132`

[6] More specifically, results presented here were calculated based on data from Historical Indexes, ©2015 Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business. Wharton Research Data Services was used in preparing this article. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

[7] The 10-year Treasury index was constructed from monthly returns from CRSP back to 1941. The data for 1926-1941 were interpolated from annual returns in "A history of interest rates," S. Homer and R. Sylla.

	Units: thousands of dollars (real)					
					Probability of Shortfall	
Strategy	$E[W_T]$	$Median[W_T]$	$std[W_T]$	CVAR (5%)	$W_T < 500$	$W_T < 700$
100% 10-yr US Treasuries	456	431	147	242	.68	.93
Cap-weighted (60 : 40)	878	780	442	312	.17	.41
Equal weighted (60 : 40)	1117	974	633	370	.08	.25

TABLE 0.1: *The investor starts with zero initial wealth at time zero, and injects \$10,000 real at* $t = 0,1,\ldots,29$ *years. The final results are given at* $T = 30$ *years. Block bootstrap resampling results based on the real CRSP stock index data (US stocks) and real 10-year US treasuries, for the period 1926:1-2015:12. 100,000 resamples were used. Cap-weighted* (60 : 40)*: refers to 60% capitalization weighted CRSP stock index, and 40% 10 year US treasuries, rebalanced yearly. Equal weighted* (60 : 40)*: refers to 60% an equal weighted CRSP stock index, and 40% 10 year US treasuries, rebalanced yearly. Total investment horizon is* $T = 30$ *years.*

Results

Our investing scenario assumes an investor who starts off with zero initial wealth, and invests \$10,000 (real) each year, for 30 years. The investment portfolio receives the cash and is rebalanced also annually. We will assume that the investor injects the cash at times t, $t = 0,1,\ldots,29$ years, and we will examine the statistics of the final portfolio value at $T = 30$ years.

We show the results obtained using 100,000 bootstrap resamples in Table 0.1. Each resample consists of a set of 360 monthly returns. We use an average blocksize of 24 months [8]. The column labelled $E[W_T]$ is the expected (average or mean) value of the final portfolio wealth W_T at time $T = 30$ years, based on all the 100,000 resamples. The $Median[W_T]$ is the value of wealth such that 50% of the paths end up below the median, and 50% end up above the median. $std[W_T]$ refers to the standard deviation of the final wealth, which is a conventional measure of the variability around the mean value. The columns labelled *Probability of Shortfall* give the probabilities that we end up with final wealth less than \$500,000 and \$700,000.

You may have seen such statistics before. Probably you have not seen the column labelled CVAR (5%). This is the Conditional Value at Risk, at the 5% level. This is just the average of the worst 5% of the outcomes. This is a measure of *tail risk*. This answers the question: if the outcomes are bad (worst 5%), what is the average of these bad cases? Or in other words, if things turn out bad, how bad can it be?

Our scenario assumes that the investor has contributed \$300,000 in constant dollars to the portfolio. A CVAR of less than \$300,000 would indicate that 5% of the time, the investor (on average) loses money. On the other hand, a large value for CVAR indicates that even in the worst 5% of the outcomes, the result is still good. So, we would like to see a large number for CVAR.

We can illustrate the idea of CVAR by examining the probability density curve, which we show in Figure 0.1, for the (60 : 40) strategy using the equal weight CRSP index. Suppose we want to know the probability that the final wealth W_T is larger than the mean value $E[W_T]$. This is given by the area under the probability density curve to the right of the green dotted line representing the mean of the final portfolio value ($E[W_T]$). By definition, the area under the entire curve will be one (i.e. it is certain that the final wealth with be between zero and infinity). The area under the

[8] "Correction to: automatic block-length selection for the dependent bootstrap," authors: Patton, Politis, White; Econometric Reviews vol 28 (2009) 372-375

curve to the left of the 5^{th} percentile line (blue dotted line), is .05. The CVAR (5%) is the average of the wealth values occuring under the orange shaded region to the left of the 5^{th} percentile line.

Note that we are focusing on the statistics of the final wealth after 30 years, as opposed to comparing rates of return. You can only spend dollars, not rates of return, so it is informative to look at the actual distribution of wealth.

One thing you can see from Figure 0.1 is that the density curve is highly skewed, with a long right tail of low probability, high portfolio values. This causes the mean value to be larger than the median value. The CVAR statistic focuses on the bad outcomes in the left tail of the distribution.

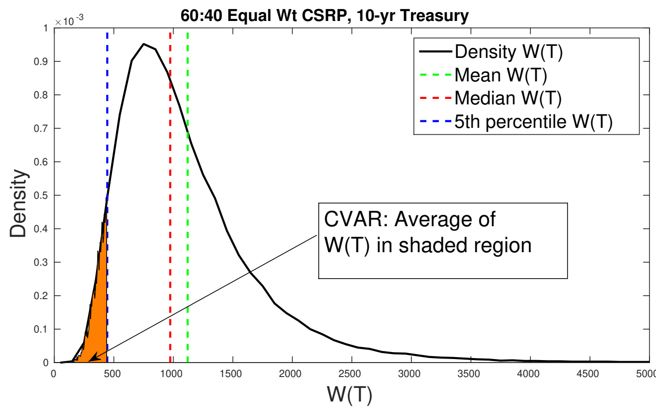

FIGURE 0.1: *Probability density of W_T, based on bootstrap resampling of historical data 1926:1-2015:12. Strategy: 60% equal weighted CRSP index and 40% 10 year US treasuries.*

The first strategy we examine is simply investing in a 10-year US treasury index, for the entire 30 years. You can see from Table 0.1 that there is a significant variability of the investment results, with a standard deviation of $147,000 compared to the expected outcome of $456,000. As well, the mean of the worst 5% of the outcomes (CVAR) is only $242,000 (real USD). The variability comes about due to the risk of investing in 10-year treasuries in times of inflation and rising interest rates.

The results for the traditional portfolio which is invested 60% in a broad cap-weighted stock index and 40% in a bond index is much better, by almost all measures. The median value of the final portfolio is $780,000 compared to only $431,000 for the all bond strategy. As well, the probability of ending up with less than $500,000 is only .17, compared with the bond portfolio (68% probability of ending up less than $500,000). Even the CVAR for the 60:40 portfolio is larger (better) than for the all bond portfolio.

Finally, the last row of Table 0.1 shows the results from bootstrapping the 60% equal weighted stock index, and 40% bond index. By all measures (except standard deviation), the equal weighted stock index beats the other strategies. In fact, I don't think that standard deviation is a very useful measure of risk here. The 5% CVAR is quite a bit larger than for the cap-weighted portfolio. Using the equal weighted index gives a probability of ending up with less than $500,000 of .08, less than half of the corresponding probability of the using the cap-weighted index.

So, consistent with previous research, this suggests that an equal weighted index beats a cap weighted index.

Shall we equal weight?

However, there is a fly in the ointment here. If you actually look at the available equal weighted ETFs, you will find that they have fairly hefty fees. In addition, all this buying and selling (when

Strategy	$E[IRR]$	$Median[IRR]$	$std[IRR]$	CVAR (5%)	5^{th} percentile
100% 10-yr US Treasuries	.022	.022	.018	-.014	-.007
Cap-weighted (60 : 40)	.054	.055	.024	.004	.014
Equal weighted (60 : 40)	.066	.066	.025	.012	.024

Table 0.2: *Real internal rates of return (IRR) for strategies described in Table 0.1. Statistics computed using bootstrap resampling of historical data.*

the index is rebalanced to equal weight) generates realized capital gains, with tax consequences.

To be more concrete here, Table 0.2 shows the statistics for the internal rate of return (IRR). The IRR is the average annualized real rate of return over the entire investing period, taking into account the timing of the contributions to the portfolio. You might have seen this on your brokerage statement, where it is commonly referred to as the *personal rate of return* or the *money weighted rate of return*. From Table 0.2, we can see that the equal weighted CRSP index strategy beats the cap-weighted strategy by $.011 - .012$, depending on whether we look at the median or the mean return. These sorts of return differences are often quoted in basis points, or 1/100 of a percent. So, the return of an equal weight (60 : 40) strategy exceeds that of the cap-weighted (60 : 40) strategy by about 110 bps (basis points or, if you want to sound cool, this is 110 *bips*).

So, how much can we afford to pay in extra fees for the equal weighted strategy? The equal weight index is 60% of the portfolio, and so contributes .60 to the total costs. Therefore, the total fees (MER plus trading expenses) being charged by a candidate equal weight ETF should be less than $110/.60 = 183$ bps, to make this worthwhile. I remind you that we are ignoring taxes here. Probably you need to hold the equal weight ETF in a non-taxable account.

You also need to check out the rebalancing frequency of the fund. I assumed monthly rebalancing for the equal weight ETF. Many of the funds only rebalance quarterly, which will cause some slippage.

So far, I have not found a broad market equal weight index ETF that I like. I am still looking.

Another note of caution. I have looked at the results over a 30 year investment horizon. For shorter periods (i.e. 5 years), the results will be dominated by noise, and it would be difficult to come to any real conclusions about strategies deviating from a cap-weighted index. [9]

Finally, there are some interesting theoretical questions about bootstrap resampling. If we are going to use this method to test investment strategies, we are really assuming that the future will look somewhat like the past. Is this a valid assumption? Well, we have used a very long time series, starting at 1926. This historical period encompasses the great depression of the 1930s, the Second World War, periods of high interest rates and inflation (1980s), periods of falling interest rates and inflation, and the financial crisis. The resampling procedure reshuffles all these periods randomly together. We might regard this as quite a stress test for an investment thesis.

An alternative would be to have a view about what market parameters are going forward (e.g. expected returns, volatilities) and then use a conventional Monte Carlo approach to project forward possible return paths. We could then use these paths to test investment strategies.

The problem with this approach is that usually the parameters are based on recent consensus economic thinking, which is almost always wrong.

Using the historical bootstrap method at least avoids making this sort of biased judgement call. I tend to trust this more.

[9] See “Volatility Lesssons,” authors: E. Fama and K. French, `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3081101`.