Stress Testing Your Plan

Insights and Limits of Historical Returns

Ronald E. Madey, CFA

January 15, 2019

WEALTHCARE
financial guidance for life

Plans are useless, but planning is indispensable.

Dwight D. Eisenhower

You have a financial plan. How do you know it will survive contact with the stock market? Are you taking enough risk to achieve your goals? Are you taking so much risk that a market downturn could impair your financial condition today? Are recent market declines undermining your confidence?

You can't know, but you can build confidence by evaluating risk and stress testing your plan. There are a number of methods available to stress test a plan. All involve using return data, whether it is historical or simulated, to assess the likelihood and magnitude of negative outcomes. Using historical returns is intuitive to most people. Let's see what insights we can gain from analyzing historical returns.

Sequential Rolling Historical Periods

The table below uses CRSP[1] monthly data for large cap stocks from 1926 to 2017 and sequential rolling periods from one year to 50 years, moving forward from 1926 one month at a time.

Rolling Periods		Percentile		
Length	# Observations	75%	82.5%	90%
1 year	1,093	0.0%	-6.0%	-12.8%
5 years	1,045	4.3%	1.3%	-1.6%
10 years	985	6.5%	4.6%	2.8%
15 years	925	6.5%	5.5%	4.5%
20 years	865	7.7%	7.2%	6.7%
25 years	805	9.1%	8.7%	7.7%
30 years	745	9.7%	9.5%	9.1%
50 years	505	9.4%	9.1%	8.5%
Table 1: Historical Rolling Return Analysis - Sequential				
Data from 1926 to 2017 for CRSP Large Cap Stocks				

The long-term average for the entire period is about 9.65%. Losses worse than minus 12.8% for one year are not on the table. For reference, the "one-in-hundred-year" 1-year loss is minus 42.1%. This is different than a peak to trough loss which may span more or less than one year.

The table shows percentile outcomes from 75% to 90%. The 90th percentile is interpreted to mean that, based on data, 90% of the time 1-year returns would be at least minus 12.8% or better. Ten percent of the time, 1-year return outcomes would be worse than minus 12.8%. For 30-year periods, the 90th percentile is 9.1%, meaning that 90% of the time 30-year returns would be at least 9.1% or higher while 10% of the 30-year periods would have annualized returns that are less than 9.1%. The other percentile columns are interpreted similarly.

At first look, if your plan can survive the first 10 years of negative/low returns at the 90th percentile, you could

[1]The Center for Research in Securities Prices

feel highly confident about the 30-year period. However, there are flaws with this analysis. Here are a few:

1. Not all monthly data is weighted the same which biases the results. For example, using the sequential rolling return method in the above table, January of 1926 and December of 2017 are used only once under each length tested. For the 1-year length, January 1927 and December 2016 are used 12 times, as is every month in between. For the 30-year length, all months between January 1956 and December 1987 are used 360 times, while months before and after are used less. This biases the returns of the longer lengths higher.

2. Because of the weighting issues noted in #1, oddities occur in the table. At the 75th percentile for the 30-year length, the return shown is above the long-term average (9.65%). This is nonsensical. Another oddity: the 50-year returns are lower than the 30-year returns. As the length increases, the returns should rise and approach the long-term average but not exceed it.

3. The observations are not independent, as data is reused repeatedly and sequentially in the rolling process. There are 92 years of history in the analysis. This means that, with a length of 30 years, there are only three possible non-overlapping periods as compared to 745 overlapping periods. A 50-year length means there is only one possible non-overlapping period as compared to 505 overlapping periods.

4. An implied assumption with the approach is that we have experienced all possible market outcomes and sequences in the historical period chosen. While almost 100 years of stock market history is a lot, it does not mean there is nothing new to experience. If we had applied the approach in 1999, we would not have captured the tech bubble bursting nor the 2008/09 financial crisis. Further, most asset classes (e.g. emerging markets, high yield bonds) have much shorter data histories making it more likely that there will be outcomes that have not been experienced yet.

5. You can't apply the approach effectively for a portfolio that is diversified using asset classes with much shorter data histories – for example, TIPS, high yield bonds, emerging markets, and cryptocurrencies.

Wrapped Rolling Historical Periods

We can solve for the first two flaws by simply wrapping the sequential rolling period analysis. In this approach, when we get to 2017 using the sequential rolling process, we wrap around back to 1926 and continue to run the sequential process until each data point in the history is used the same number of times. Here are the results:

Rolling Periods		Percentile		
Length	# Observations	75%	82.5%	90%
1 year	1,104	0.2%	-5.7%	-12.7%
5 years	1,104	4.7%	1.7%	-1.4%
10 years	1,104	6.0%	4.1%	2.5%
15 years	1,104	6.3%	5.4%	4.5%
20 years	1,104	7.0%	6.3%	5.5%
25 years	1,104	7.6%	6.9%	6.2%
30 years	1,104	8.2%	7.6%	6.3%
50 years	1,104	8.6%	8.4%	8.1%
Table 2: Historical Rolling Return Analysis with Wrapping				
Data from 1926 to 2017 for CRSP Large Cap Stocks				

As you can see, the oddities of the first table are eliminated. None of the returns in the table are above the long-term average, the 50-year length has the highest returns and, more generally, as length increases, returns rise toward the long-term average but do not exceed it.

Sequential versus Wrapped Rolling Historical Periods

The next table shows the difference in returns between Table 1 and Table 2.

Rolling Periods		Percentile		
Length	# Observations	75%	82.5%	90%
1 year	(11)	-0.2%	-0.4%	-0.1%
5 years	(59)	-0.4%	-0.4%	-0.2%
10 years	(119)	0.4%	0.5%	0.3%
15 years	(179)	0.2%	0.2%	0.0%
20 years	(239)	0.7%	0.9%	1.1%
25 years	(299)	1.5%	1.8%	1.5%
30 years	(359)	1.5%	1.9%	2.8%
50 years	(599)	0.8%	0.7%	0.4%
Table 3: Sequential less Wrapped Historical Rolling Returns				
Data from 1926 to 2017 for CRSP Large Cap Stocks				

The number of observations, or sample size, gets significantly smaller for sequential method of Table 1 relative to the wrapped method of Table 2 as period length increases. For the 20-year length and longer, the overestimation of returns for the sequential method is significant – peaking at 2.8% for the 90th percentile with a 30-year length.

When considering Table 2, it is worth mentioning that most people are not 100% U.S. stocks. If you are 50% invested in stocks, then the early year losses are muted by half. What your portfolio actually does depends on the asset classes and weights of what you actually own.

The takeaway from the above analysis is that if you use historical sequential rolling period analysis, chances are you will have more confidence in future market returns and your plan than you should. If you utilize historical wrapped rolling period analysis you will be less likely to have an overly optimistic view of longer term returns for U.S. stocks. However, you can't really stress test your plan using a diversified portfolio with this method because data histories for some asset classes are short. So, unless you own 100% U.S. stocks and you are willing to accept the other flaws of the analysis listed in items 3 through 5 above you need a better method.

Bootstrapping

Bootstrapping is a statistical technique that uses random sampling with replacement. Wealthcare uses this technique with its *Wealth Simulator*. Under this approach, monthly returns are selected at random from the data set. A given monthly return can be used multiple times or not at all. The approach addresses the first two of the flaws listed above and part of the third. The returns for the various lengths/horizons now have statistical independence tied to the random process, but we are still captive to the historical data set and the data is repeatedly used. The larger the data set and the shorter the horizon, the less this is a problem. Using this approach for an asset class such as large cap stocks with 92 years of monthly history is clearly more defensible than using it for TIPS (Treasury Inflation Protected Securities) with just 20 years of history. Also, to the extent that there is serial correlation, positive or negative (mean reversion) in the data, the information is lost with bootstrapping as compared to rolling period analyses.

Table 4 show the results from the bootstrapping approach and Table 5 compares it to the wrapped historical rolling period approach. Results for the Table 4 are based on 100,000 trials using the bootstrapping method.

Horizon/Length	Percentile		
	75%	82.5%	90%
1 year	0.2%	-5.4%	-12.6%
5 years	4.1%	1.5%	-2.1%
10 years	5.5%	3.6%	1.2%
15 years	6.1%	4.6%	2.7%
20 years	6.6%	5.3%	3.6%
25 years	6.9%	5.7%	4.2%
30 years	7.1%	6.1%	4.7%
50 years	7.7%	6.9%	5.8%

Table 4: Wealth Simulator/Bootstrapping
Data from 1926 to 2017 for CRSP Large Cap Stocks

Table 4 has the desirable properties of returns rising with horizon and absence of the oddities that we observed in Table 1. However, Table 5 shows that except for the 1-year horizon, the bootstrapped returns are systematically lower than the wrapped historical returns and significantly so as percentile and horizon increase. This suggests that there is information in the wrapped historical returns sequencing that is lost through the bootstrapping process.

Horizon/Length	Percentile		
	75%	82.5%	90%
1 year	0.0%	0.2%	0.1%
5 years	-0.6%	-0.2%	-0.7%
10 years	-0.6%	-0.5%	-1.3%
15 years	-0.2%	-0.7%	-1.8%
20 years	-0.4%	-1.0%	-1.9%
25 years	-0.7%	-1.2%	-2.0%
30 years	-1.1%	-1.5%	-1.6%
50 years	-1.0%	-1.5%	-2.3%

Table 5: Bootstrap less Wrapped Rolling Periods
Data from 1926 to 2017 for CRSP Large Cap Stocks

The bootstrapping process assumes no correlation, positive or negative, from one month or year to the next. The sequence of returns is not preserved in the bootstrapping process, but is with the wrapped historical returns. The wrapped historical returns exhibit mean reversion at longer horizons. Said another way, after a longer period of below average returns, the likelihood of a several year period of above average returns is greater than random, resulting in lower long-term and tail returns for the bootstrapping method as it randomizes the data. While the wrapping process preserves the sequence of returns, it is worth reminding ourselves of certain flaws with the wrapping process – there is a discontinuity in the sequence for periods that are wrapped and there is only one possible fully independent observation for the 50-year horizon, just three for the 25-year horizon and nine for the 10-year horizon given our data set.

Key takeaway from the bootstrapping analysis: If your plan can survive the bootstrapping stress test for large cap stocks, you will likely have confidence that you will not run out of money. However, you may also unduly sacrifice your current lifestyle and life goals while ending up with a much larger estate for your heirs than you intended, as the longer term downside returns under the bootstrapping method are significantly lower than what has occurred historically.

Monte Carlo Simulation

The most commonly accepted method to address all the flaws listed above is Monte Carlo simulation. This is the method Wealthcare uses to provide the patented Comfort Zone® to clients. In Monte Carlo simulation, return, risk, and correlation assumptions (Capital Market Assumptions or "CMA") are developed for every asset class in your portfolio. Based on these assumptions, thousands of return series over the life of your plan can be generated based on your portfolio and used to stress test your plan. The quality of the stress test depends on the quality of the CMA.

A full discussion of CMA is a topic for another paper. For the purpose of this paper, we built the expected return and risk assumptions based on the CRSP data from 1926 to 2017, then simulated returns based on the lognormal distribution. Here are the results:

Horizon/Length	Percentile		
	75%	82.5%	90%
1 year	-1.5%	-5.5%	-10.6%
5 years	4.5%	2.6%	0.1%
10 years	6.0%	4.6%	2.8%
15 years	6.7%	5.5%	4.0%
20 years	7.1%	6.1%	4.8%
25 years	7.4%	6.5%	5.3%
30 years	7.6%	6.7%	5.7%
50 years	8.0%	7.4%	6.6%

Table 6: Monte Carlo Simulation

Horizon/Length	Percentile		
	75%	82.5%	90%
1 year	-1.7%	0.1%	2.1%
5 years	-0.2%	0.9%	1.5%
10 years	0.0%	0.5%	0.3%
15 years	0.4%	0.2%	-0.4%
20 years	0.1%	-0.2%	-0.7%
25 years	-0.2%	-0.4%	-0.9%
30 years	-0.6%	-0.9%	-0.6%
50 years	-0.6%	-1.0%	-1.6%

Table 7: Monte Carlo less Wrapped Rolling Periods

Looking at Table 6, the patterns make sense. Returns rise as horizon increases without the oddities of Table 1. Table 7 shows the difference between the Monte Carlo results and the wrapped historical rolling periods. At the 90th percentile, the 1-year return difference of 2.1% stands out. This is likely related to shorter-term positive serial correlation and the "fat-tail" phenomena documented in the literature. Fat tails refer to the evidence that shorter term returns have greater extremes than are expected with the lognormal distribution. However, as you can see from Table 6, the results reverse as the horizon increases, partly because longer-term mean reversion and the flaws of the rolling return process – lack of independence/repeated use of data – have a greater impact as horizon increases. Overall, the Monte Carlo approach produces returns close to the wrapped historical return approach, particularly in the 10 to 25 year horizons. Yet, it solves for challenges that the wrapped historical return approach does not and cannot.

Summary & Conclusions

We examined four different ways to utilize historical data to stress test a financial plan using 92 years of monthly stock returns. They are:

1. Sequential rolling historical returns
2. Wrapped rolling historical returns
3. Bootstrapped historical returns
4. Monte Carlo simulation

Only the Monte Carlo simulation approach solves all five of the flaws listed above for the simple sequential rolling return approach. With Monte Carlo simulation, returns are statistically independent; there is not a dependency on a specific historical data period; asset classes with short data histories are easily included in the framework, enabling the stress testing of diversified portfolios holding newer asset classes; within the framework of the lognormal distribution, all market scenarios are possible, even ones we have not experienced yet. While almost 100 years of stock market history is a lot, it does not mean there is nothing new to experience. If we had applied a historical approach in 1999, we would not have captured the tech bubble bursting nor the 2008/09 financial crisis.

While Monte Carlo simulation is not a panacea, it is an effective risk management tool. Central to its successful use are well-researched and plausible Capital Market Assumptions. To build effective long-term CMA, historical data is used in conjunction with economic relationships and financial market theory.

Outside the scope of this paper is another advantage that Monte Carlo simulation offers relative to the alternative approaches – the ability to capture current conditions, such as the current level of inflation, the current level of interest rates, and equity market valuation to adjust the early returns of the simulation. For example, the first ten years of the simulation can use expert forecasts for CMA and then revert to the long-term CMA for the rest of the simulation for the stress test of a 30, 40 or 50 year plan.

Sources in include The Center for Research in Securities Prices, Wealthcare.

Thanks to Chris Philipoom for his research support in producing this paper.

WEALTHCARE
financial guidance for life

Two James Center 1021, East Cary Street, Suite 1120 | RICHMOND, VA 23219 | 804.644.4711

WEALTHCARECAPITAL.COM