Retirement Risk Assessment: History Is Not a Guide
September 24, 2018
by Joe Tomlinson

In working with clients to develop retirement spending strategies, advisors need to consider worst-case potential outcomes due to poor investment performance. One way involves using historical returns for bonds and stocks and focusing on the particular periods that would have produced the worst outcomes. Another way uses forward-looking estimates of average future returns and Monte Carlo simulations to generate a range of potential outcomes. I’ll develop new measures to compare those two approaches and demonstrate why I strongly favor the forward-looking approach.

Historical returns

The most noteworthy financial planning research study of all time is Bill Bengen’s classic *Determining Withdrawal Rates Using Historical Data,* published in the *Journal of Financial Planning* in 1994. Bengen demonstrated that, for all historical 30-year periods going back to 1926, a portfolio of at least 50% stocks would have supported withdrawals of 4% of the initial portfolio, increasing with inflation each year – hence the “4% rule.” For the analysis that follows I’ll be assuming withdrawals based on the 4% rule in order to compare downside risk based on historical returns versus forward-looking return estimates. I won’t evaluate the viability of the 4% rule as such, but instead use it as a basis to compare retirement outcomes.

When Bengen published his study 25 years ago, 30 years was a conservative estimate of the potential length of retirement. But longevity has improved, particularly for the upscale individuals typical of advisory clients. I developed an estimate for a 65-year old couple where the life expectancies are 88 for the husband and 90 for the wife. The expected age at death for the last member of the couple is very close to 95. So a 30-year retirement has evolved into more of an average estimate rather than a conservative one, particularly for client couples.

Using these updated longevity estimates, I did a new version of the Bengen analysis based on the Ibbotson data for stock and bond returns (large company stocks and intermediate-term government bonds) and a 60/40 stock/bond portfolio. I did this for cohorts of retiring 65-year-old couples, with each cohort associated with a retirement start year 1926-1978. I stopped at 1978 so that I would have at least 40 years of actual data for all of the cohorts. Instead of using a 30-year retirement duration, I made the longevity variable to reflect what a population of couples might actually experience – some with short lives, others with longer lives – and measured retirement duration based on the last to die for each such couple. A given couple in a particular retirement-year cohort would experience the same year-by-year investment returns as other members of that cohort, but longevity would be variable for the couples in a cohort.

I came up with a new measure for average stock and bond returns – for a given retirement start year, the average returns reflect the forward-looking average arithmetic real returns over retirement for that particular cohort. Because members of the cohort are assumed to die off over time, these returns are effectively survival-weighted – the early retirement years get more weight. The chart below shows the average returns by retirement start year.
What we see is that over the past 90 years average retirement returns for stocks and bonds have tended to offset, mitigating retirement risk. For example, the 1933 through 1946 retirement cohorts all experienced negative average real bond returns, and these returns did not climb above 2.00% until 1959. But for these particular cohorts, stock portfolios did particularly well. Even when stock returns were at their lows for the late 1950s through 1960s cohorts, average bond returns over the course of retirement were making a comeback, which mitigated the poor stock returns. The full 90 years, 1926-2017 produced very favorable results to support retirement portfolios – both the level of returns and the way stocks and bonds have offset.

We can add emphasis to the stock/bond offsets by examining correlation coefficients. In terms of the real returns experienced by retirement-year cohorts, the stock/bond correlation coefficient was negative 59% – further illustrating that when retirement cohorts do poorly with bonds, they tend to do well with stocks and vice versa. By contrast the correlation coefficient calculated in the traditional way for annual stock and bond returns was a positive 9% – a completely different picture than for the cohorts.

**The failure cases**

I’ll now see what information we can obtain about retirement risk by focusing on the particular retirement-year cohorts that experienced plan failure, i.e. where members of the cohort outlived their savings. Because I am using historical year-by-year stock and bond returns, all couples in a retirement-year cohort experience the same investment returns and take the same withdrawals, so portfolios for each cohort will be depleted at the same time. The portfolio-failure percent is a measure of the proportion of couples in a cohort who outlive this portfolio depletion date.

Of the 53 retirement-year cohorts, 44 made it completely though retirement without the portfolio being depleted, no matter how long the oldest member lived – up to 120 in a very few cases. The chart below shows outcomes for the nine cases where portfolios were depleted.
As Bengen noted in his original study, the 1960s would have been a particularly challenging time to retire and initiate inflation-adjusted withdrawals in accordance with the 4% rule. The worst year would have been 1966. Among those couples retiring then, 60.9% would have outlived their portfolios, which would have been depleted at age 93. That is the only cohort whose portfolio would not have lasted for at least 30 years to age 95. So switching from 30-year retirements to variable longevity raises the failures from 1 to 9. The bottom line of the chart shows the expected experience for all 53 failing and non-failing cohorts combined.

In the first three columns I show investment returns for the particular failing cases identified by retirement-start year. These correspond to the graphical version presented in the first chart for all the cohorts. These are the arithmetic average yearly real returns for stocks and bonds (plus the 60/40 weighted average) experienced by each of the retirement-year cohorts. As expected, the weighted average returns for the cohorts with failures are significantly lower than for the overall population shown in the last line of the chart.

The 1966 cohort, which has the worst failure rate, does not have the worst average return. This is an indication that sequence risk played a role — poor real returns (including effects of high inflation) early in retirement. The chart also shows median bequests for each of the failing cohorts and, as might be expected, these vary directly with the failure percentages and are significantly below the overall median.

Causes for concern

There are few, if any, respected economists, investment managers or thought leaders in the financial planning community who expect future returns for stocks and bonds to match historical averages. For bonds, returns in the early retirement years will mostly be determined by bond yields at the time of retirement, and it is returns in the early years that have the biggest impact on retirement outcomes. Yields on 10-year Treasury inflation-protected securities (TIPS) are 0.86% as of mid-September 2018. By comparison, the historical real yields on intermediate-term government bonds from Ibbotson have averaged 2.26%. It is virtually impossible to conceive of an economic scenario that would drive TIPS yields far enough into negative territory to boost average annual real bond returns to anything approaching the historical average.

Future returns on stocks are not as locked-in as bond returns, but dividend growth models projecting future stock returns typically assume 2% dividends growing at a real rate of at most 3%, so it’s hard to come up with an estimate above 5%. Those who pay attention to stock valuation measures, such as the popular CAPE measure developed by Yale economist Robert Shiller, express concern that today’s CAPE level is about double the historical average. In this recent CNBC interview summary, Shiller forecasts a 10-year forward return for the stock market of around 2.6%. His views have been counterbalanced by Jeremy Siegel who has consistently held a much more optimistic view for stocks. But Siegel is quoted in the summary as saying that stocks may be overvalued, but bonds are even more overvalued. Thus we could see subpar future returns from both stocks and bonds simultaneously – a much grimmer scenario than the compensating returns we have experienced historically.

It’s impossible to pin an exact number for future stock returns, but in the chart below I compare performance measures based on historical performance with Monte Carlo simulations of performance based on arithmetic average returns for stocks of 5% and 3%, and today’s TIPS yield as a forecast bond return. It’s a bit unnerving that the forward-looking weighted average returns are significantly below not only historical averages, but even historical returns for the worst
cases shown on the prior chart.

<table>
<thead>
<tr>
<th>Source of returns</th>
<th>Real Stock Return</th>
<th>Real Bond Return</th>
<th>Weighted Average</th>
<th>Median Bequest</th>
<th>Failure Percent</th>
<th>Average Failure Age</th>
<th>5th Percentile Failure Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical 1926 - 2017</td>
<td>9.04%</td>
<td>2.26%</td>
<td>6.33%</td>
<td>$1,842,248</td>
<td>5.7%</td>
<td>NA</td>
<td>94</td>
</tr>
<tr>
<td>Forecast stock return 5%</td>
<td>5.00%</td>
<td>0.85%</td>
<td>3.34%</td>
<td>$274,765</td>
<td>31.5%</td>
<td>102</td>
<td>84</td>
</tr>
<tr>
<td>Forecast stock return 3%</td>
<td>3.00%</td>
<td>0.85%</td>
<td>2.14%</td>
<td>$46,919</td>
<td>46.6%</td>
<td>94</td>
<td>82</td>
</tr>
</tbody>
</table>

Source: Author’s calculations

We see that the forward-looking failure percentages increase dramatically. I also calculated failure ages, when funds would be depleted for the 65-year old couple taking 4% rule withdrawals. If we use 5th percentile failure ages as a risk measure, we see that historically we would fall just short of 30 years, very similar to the worst failure ages in the prior chart. But, with lower return forecasts, we don’t even make 20 years – quite a different risk outlook. This difference in risk assessment would apply regardless of the retirement withdrawal strategy being utilized – fixed withdrawals where the impact is on the failure age, or withdrawals that vary with underlying investment performance where the impact is on the level of withdrawals.

Final word

Stocks have delivered real returns averaging close to 14% per year for the past nine years, so it may be difficult to imagine a future scenario where returns drop down to the 3% to 5% range. However, stocks can only deliver what comes from dividends and dividend growth plus increases on P/E ratios, and, with P/Es at elevated levels, the direction is more likely down than up.

Bond returns for intermediate and long maturities are largely locked in at current yield levels. For retirement risk assessment, worst-case analysis based on historical returns paints too rosy a picture. Instead, Monte Carlo simulations based on realistic average return forecasts, with a focus on results at the lower percentiles, presents a much more accurate indication of risk.

Joe Tomlinson is an actuary and financial planner, and his work mostly focuses on research related to retirement planning. He previously ran Tomlinson Financial Planning, LLC in Greenville, Maine, but now resides in West Yorkshire in the UK.