**Goal-Based Investing: A Simulation Approach**

**Paper Scope**

Although this paper only deals with a single asset, this general approach is compatible with modelling the relationships between multiple assets. Although a full discussion of this topic is outside of the scope of this paper, the scenarios may be linked together using a Bayesian Network. Shenoy has already shown that Bayesian Networks are compatible with MPT. It is the intention of the author to delve deeply into this topic in a future paper.

The proposed approach also may be extended to incorporate taxes and tax rules (e.g. tax loss harvesting) into the analysis. As discussed by Brunel, this is critical. However, this topic complicates the analysis and thus is beyond the scope of this paper.

**An Advantage of Goal-Based Investing over Modern Portfolio Theory**

Suppose an investor has a choice today between one of two weighted coins each with payout occurring in one year.

1. The coin has a .2 probability of showing heads. On heads, the payout is $25,000; whereas, tails pays zero.
2. The coin has a .5 probability of showing heads. On heads, the payout is $20,000; whereas, the investor must pay $5,000 if the coin shows tails.

Since these types of coin flips are mathematically equivalent to a transformed Bernoulli Distribution, their Expectation and Variance are analytically calculated. Suppose is a random variable such it conforms to a Bernoulli Distribution. Then, it has the following properties.

Suppose a linear transformation is applied to using a scaling constant and shift constant such that . Using the properties of Expectation and Variance, the following is true.

The scenarios above have different parameters for , , and .

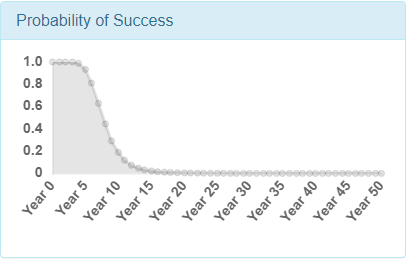
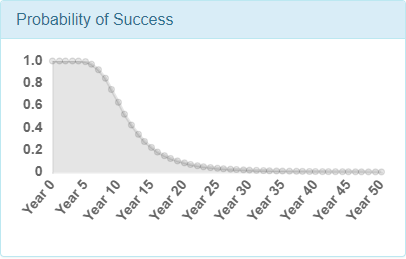
Therefore, the expectations and variances are as follows

Since both of these scenarios have the same Expectations and Variance, they are equivalent under Modern Portfolio Theory. However, they are not necessarily equivalent from the perspective of Goal-Based Investing. An investor with a required cash flow of $22000 in one year has a 20% chance of reaching his goal with Scenario 1 and 0% with two. Therefore, Scenario 2 is certainly riskier. If the required cash flow is $15000 instead, the investor has a 50% chance of matching it with Scenario 2 and 20% with Scenario 1. Therefore, Scenario 1 is riskier. Consequently, Goal-Based Investing is more discerning in terms of risk than MPT. In particular, it is sensitive to assets whose distributions are asymmetrical and fat-tailed.

**Client Narrative**

Rose is a 70 year old widow who recently received a lump sum distribution of $400,000 from a law suit. She receives income from a life insurance policy and social security, which covers her basic expenses. Her goal for the discretionary portfolio is to fund private school for her grandson. The best private school in the area costs $40,000 in inflation-adjusted dollars each year, and he has 10 years before he graduates.

Rose consults with her financial advisor, Meredith, on how to allocate funds. In addition, she mentions that she is only comfortable investing in stocks and annuities. Although Meredith disagrees with such a severe constraint on the investable universe, she agrees to determine the probability of Rose meeting her goal. To start, Meredith simulates the performance of an equal allocation of $200,000 in both stocks and an annuity. Her software subtracts $40,000 from the invested amount for each simulated year. If the balance ever turns negative, then that simulation is a failure. If not, it is considered to be a success. Meredith generates 10,000 such simulations and only 19% meet the goal. Rose suggests testing different allocations, and Meredith creates Graph 1.

***Graph 1: Probability of Goal Attainment for $200,000 and $400,000 invested in Stocks based on a budget of $400,000, initial cash outflow of $40,000, and 2% inflation. Annuity assumed to pay 6% of principal. Note: investing in annuity alone would never meet required outflows and is not shown***

Investing in 100% results in the highest probability of success of x%, but this seems too low to Rose. Perhaps, she could find a less expensive school, or only send her grandson to the best school for middle school and high school.

**Investment Forecast Narrative**

The investment simulations in the client narrative require a long-term forecast for stocks. Meredith gathered daily S&P 500 returns from 12/1/2007 to 12/1/2017 and resampled these with replacement 1000 times for 252 to create a baseline forecast. She then alters this baseline to match current conditions according to her judgment. She determines the worst and best annual outcomes for stocks might be -50% and 50% returns. On average, she expects stocks to return about 5% and assigns this to be the most likely outcome. Lastly, the historical volatility is higher than she expects going forward, and her forecast reflects this view. Lastly, she determines that a slightly negative skewness, or asymmetry in the risk and reward, is appropriate and slightly positive excess kurtosis, or “fat-tails,” is also plausible. From an alternate view, she determines that 80% of returns should fall between Y% and Z%. Table 1 summarizes the results:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Minimum | 10th Percentile | Median | 90th Percentile | Maximum |
| -48.7% | -16.3% | 8.9% | 40.0% | 120.3% |

***Table 1: Annual summary statistics from resampled gains and losses of the SPDR S&P 500 ETF from 12/1/2007 through 12/1/2017 (Daily Gains and Losses from Yahoo Finance)***

Meredith’s software draws a probability distribution based on her scenarios. 

***Graph 3: Estimated annual long-term forecast for the stock market assuming a current price of 100. The software generates the above graphs based on the input scenarios and interpolating straight lines between each scenario.***

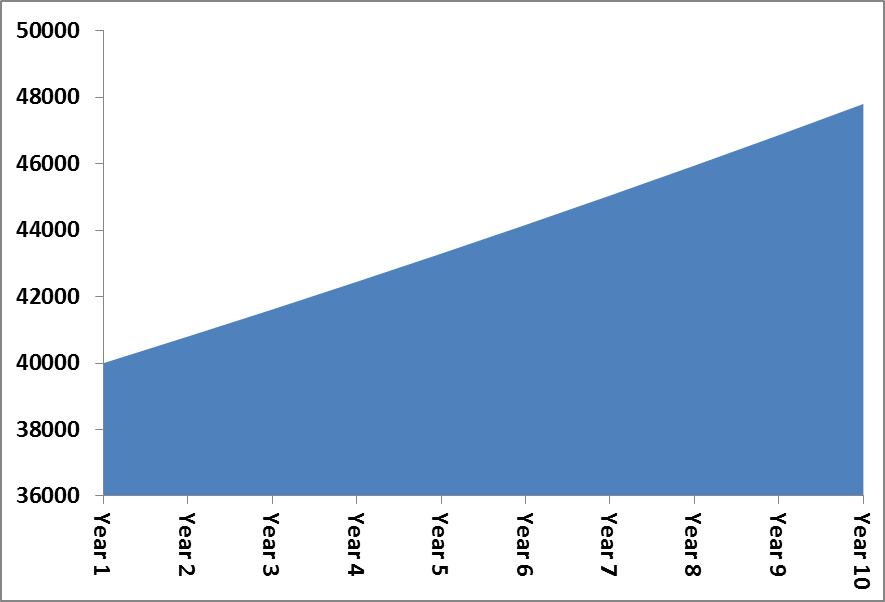
The software assumes the Probability Density Function is piecewise linear between the scenarios. It produces simulations by drawing from a Uniform Random distribution and using the Inverse Cumulative Density function of the distribution to map these simulated probabilities to returns. To illustrate, simulated probability of 0, .5, and 1 map to 50, 105, and 150. These simulated returns are aggregated into 10 year groups and matched against the cash flows. This results in the simulated probabilities of goal attainment.

**Methods**

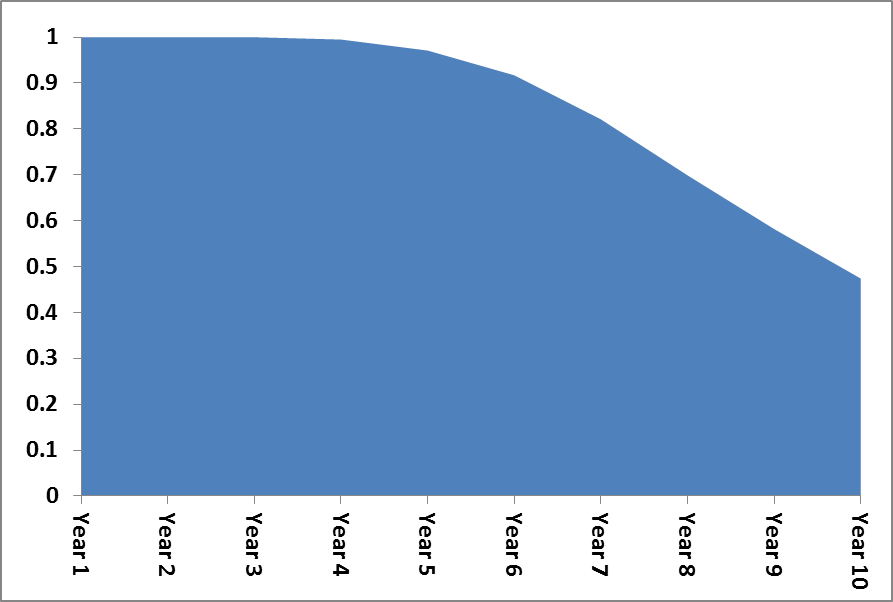
The above narrative illustrates holistically the steps required to calculate the probability of goal attainment using simulations. This section dissects this narrative into its components. Appendix A exhibits the formal mathematics underlying the required calculations. Below, the steps are explained without mathematics and graphically.

*Cash Flow Projection*

First a goal must be phrased in terms of the cash outflows. The time period of interest and intervals are required. The above narrative specifies 10 years of annual cash outflow . To ease calculation, an initial outlay is assumed for the first interval, and an assumed rate of inflation yields the remaining cash outlays. Figure BBB depicts the results from these steps.



Next, the portfolio allocated to the goal is simulated. This simulation could consist of resampled historical data or an assumed distribution, such as the Normal distribution, with the parameters selected based on the judgment of the forecaster. The section below advocates for a method that combines historical data, to the extent it is available or applicable, with forecaster judgment. Regardless of the method chosen, the cash outflows are subtracted from the portfolio value at each interval. If the portfolio value ever falls below zero, then the simulation is a failure. Otherwise, it is counted as a success. The probability of successful goal attainment is calculated as the number of successful trials divided by the total number of simulations. The probability of goal attainment is cumulative; consequently, the probability of attaining a goal in 10 years is always less than the associated probability for five years. Figure CCC illustrates the typical pattern of declining probability of attainment over time.

****

**Scenario-Based Forecasting**

*Estimating the Normal Distribution with Scenarios*

A key attribute of Goal-Based Investing is that it seeks to explain asset allocation intelligibly. This section suggests a method for forecasting portfolio and security performance and risk intuitively. Rather than assuming abstract terms such as “risk” and “reward” are parameters of a probability distribution, narratives are mapped to scenarios and these scenarios are transformed into a probability distribution. Narratives in this context are expressed in plain language, such as this simplistic example:

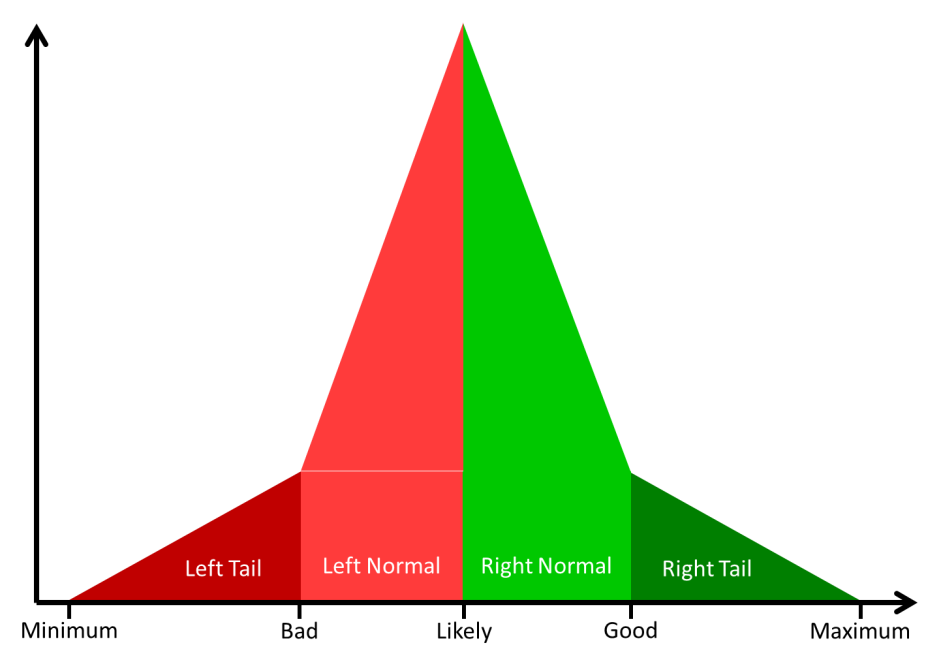
“In the past 50 years, stocks have lost at most 40% in a year. Going forward, I expect the worst possible outcome to be slightly better than the past. Therefore, 35% sounds reasonable.”

Mapped to a scenario, this becomes: “The Minimum return for stocks is 35%.”

These mappings are performed for the Maximum and Most Likely cases as well. Lastly, a narrative is constructed such as the following:

“In the past 50 years, stocks have returned between -15% and 20% 80% of the time. I suspect the future to be similar.”

This narrative would map to “Bad” and “Good” outcomes of -15% and 20%. In addition, “80%” establishes the convention that the “left-tail” and “right-tail” of the probability distribution each have probabilities of 10%. Figure DDD shows a generic Probability Density Function with labels for each of the previously defined terms.



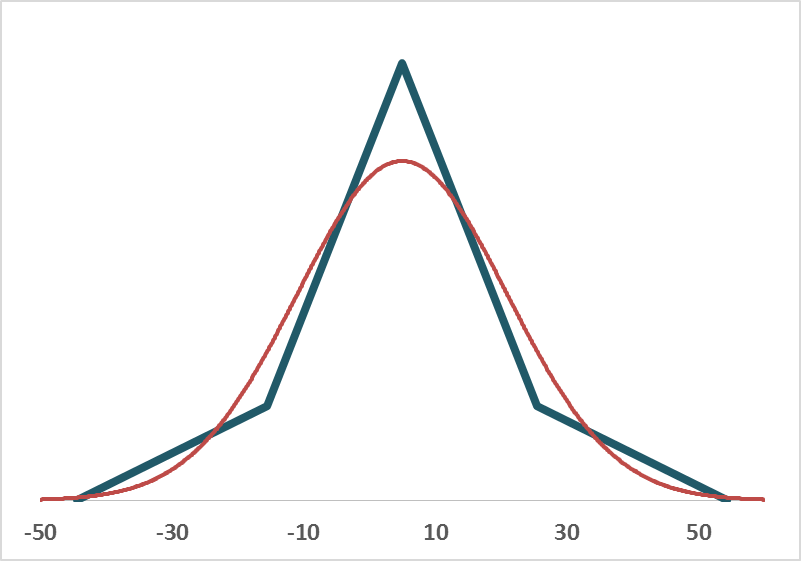
With these narratives, the investor will understand the assumptions and analysis underlying the investment forecasts required to estimate the probability of reaching a goal. As with the cliché car analogy, the investor can recognize these narratives control the calculation like a steering wheel turns a car without knowing how the underlying math works or the steering wheel manipulates the tires.

*Transforming Scenarios to a Probability Distribution*

The choice of five scenarios (Minimum, Bad, Likely, Good, and Maximum) is heuristic but not entirely arbitrary. Descriptive statistics are commonly used to characterize the first four central moments (Mean, Standard Deviation, Skewness, and Kurtosis). Five scenarios means five parameters correspond to four equations for the central moments. Therefore, five is one more degree of freedom than is required to specify the first four moments.

One additional assumption is required to “draw” the PDF from the scenarios. This is the functional form of the segments between scenarios is assumed to be linear. In other words, the Probability Density Function is assumed to be piecewise linear with the scenarios defining the ends of the pieces. Conceptually, this is analogous to estimating an integral with the Trapezoid Rule; although, the intervals in this case are uneven. Appendix B demonstrates the full specification of the PDF mathematically.

Practical results justify this assumption. Suppose a piecewise linear PDF based on scenarios is meant to approximate the normal distribution with mean of 5 and standard deviation of 16. If the Minimum and Maximum scenarios are chosen to be at the .1 and .999 percentiles of the normal distribution, and the Likely scenario is the same as the mean, then Figure AAA shows the estimated versus actual PDF.



Visually, the scenario-based distribution approximates the normal distribution closely. In addition, the central moments of the two distributions are similar as shown in Table XXX. The moments for the scenario-based distribution are calculated analytically using the equations in Appendix C.

|  |  |  |
| --- | --- | --- |
| Statistic | Normal | Estimate |
| Mean | 5 | 5 |
| Stdev | 16 | 16.33 |
| Skew | 0 | 0 |
| Excess Kurt | 0 | .33 |

The ability of the piecewise linear estimate to closely match other distributions both visually and statistically justifies its use in this context. However, it is not meant to be mathematically optimal, and better approaches may exist. This is left as a topic for further research.

As indicated in the table, the piecewise PDF is slightly fat-tailed (Excess Kurtosis greater than 0). In general, the scenarios can create asymmetrical and “fat-tailed” distributions. As a result, forecasts for assets with extreme outcomes such as venture capital and catastrophe bonds are possible. As discussed earlier, Goal-Based Investing defines risk such that the Probability of Goal Achievement is sensitive to extreme and asymmetrical outcomes. Therefore, this simulation approach capitalizes on one of the primary advantages of GBI over MPT.

*Creating Baseline Scenarios from Historical Data*

The previous subsection shows that a probability distribution based on scenarios can closely resemble a normal distribution that has been defined based on judgment. This subsection describes how a base forecast can be constructed using historical data.

If the forecast horizon is short (e.g. a month), sufficient historical returns exist, and the collected returns are all relevant to the present forecast, then descriptive statistics can be calculated directly from the historical data. For example, the Maximum base forecast would be the maximum from the sample. Standard estimators can be used to calculate the central moments, such as the mean. The forecaster can verify the reasonableness of these quantities and adjust the forecast accordingly.

However, it is most likely more practical to quarterly or annual forecasts. In this case, sufficient historical annual returns may be difficult or impossible to obtain or these quantities are deemed irrelevant to the present period. For example, are stock market returns from 30 years ago still relevant in determining such quantities as volatility today? If the answer is “no,” then another method of obtaining sufficient data for the forecast may be useful. For example, the historical daily returns for the last business cycle may be yield a sufficient number of relevant observations to the present period. Appendix E describes a method for “bootstrapping” estimated annual returns from daily observations. This requires resampling a certain number of daily returns to mimic a year of performance (e.g. 252 simulated days), compounding them to create an annual estimate, and then repeating the resampling until a representative sample of annual estimates is obtained. Lastly, the same statistical methods can be applied to these estimated annual returns as with actual historical data.

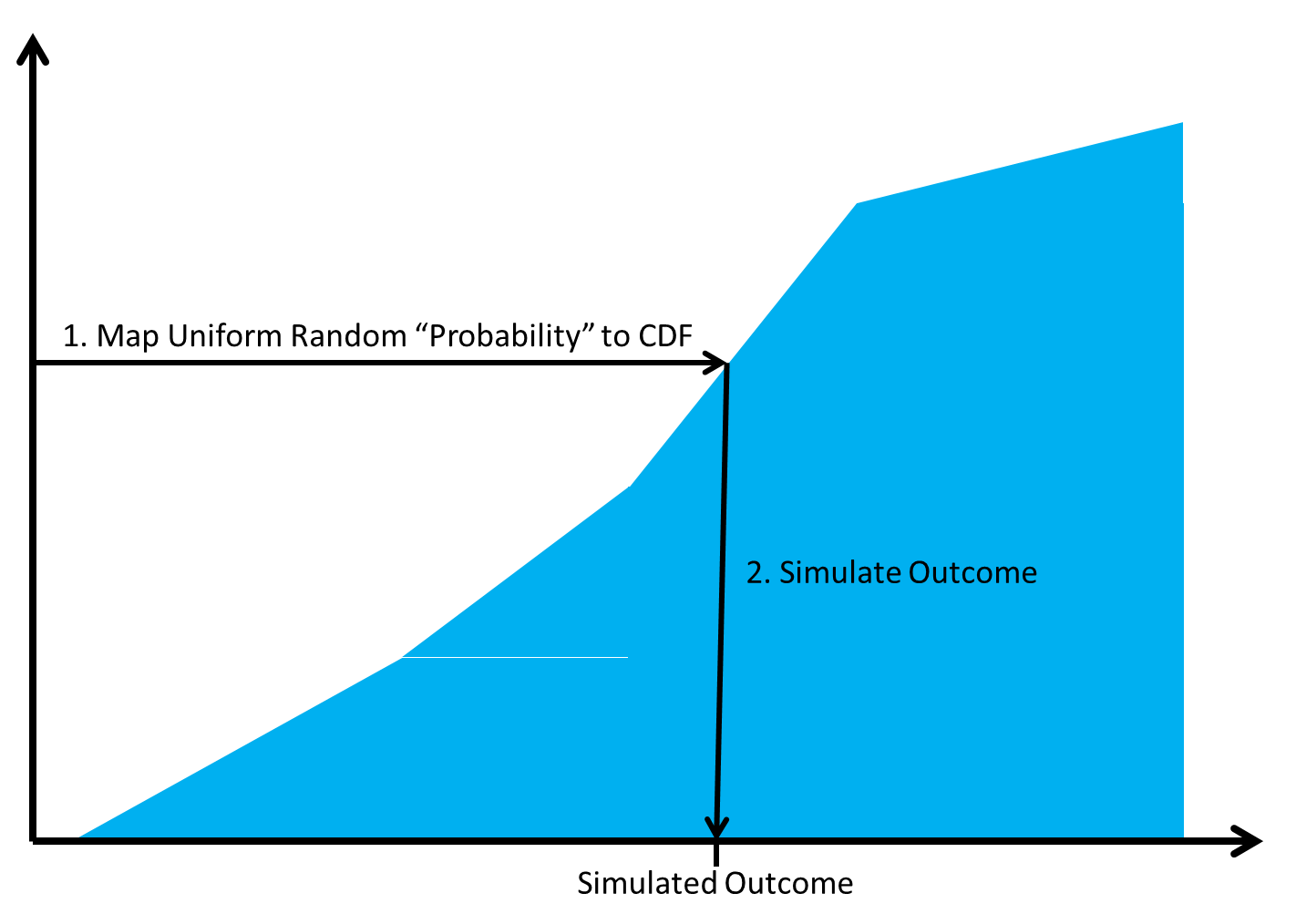
*Simulating Investment Performance*

Thus far, this section has discussed methods for mapping historical data and judgment to a Probability Density Function using scenarios. This subsection describes how the resulting Cumulative Density Function maps Uniformly Random “probabilities” to simulated investment returns.

As shown in Figure DDD, two steps are required to simulate investment returns.

1. A simulated probability is drawn from the Uniform Random distribution. Since the Y axis in Figure DDD represents probability, a horizontal line between the axis and the CDF maps the simulated probability to the corresponding point on the CDF.
2. To determine the corresponding simulated return, a vertical line is drawn from the CDF to the X-axis, which represents investment returns. The X value of the intersection is the simulated return.

Mathematically, the above simulation requires the Inverse Cumulative Density Function. Appendix E derives the relevant equations for Scenario-Based Investment Simulation.



**Framework Objectives**

**Investors:**

1. Accurately assess realism of financial goals
2. Communicate assessments in understandable terms
3. Trace assessment results back to assumptions transparently
4. Lower expenses through spreading forecast costs across many individuals
5. Prevent Forecasters from gaming forecast accuracy and quality metrics

**Forecasters:**

1. Accurately represent views
2. Communicate forecasts intuitively
3. Measure forecast quality based on accuracy and precision
4. Increase revenue through ability to leverage single forecast across many investors

**Technology:**

1. Transform forecasts into assessments of investor goals
2. Perform calculations within a reasonable time period
3. Distribute any forecasts to estimate any investor goals to lower costs
4. Build trust through understandable, open-source code
5. Model portfolios with any type of asset
6. Capture investments whose payouts can be asymmetrical and extreme (fat-tails)
7. Rely on human judgement as well as historical data
8. Minimize underlying assumptions regarding investment risk and investor risk tolerance
9. Encompass existing portfolio metrics associated with Modern Portfolio Theory and extend them
10. Measure forecaster ability objectively to enable easy comparisons

**Appendix A: Estimating the Probability of Achieving an Investment Goal**

Suppose the cash outflows required to reach a goal consists of a finite sequence of positive real numbers with at least one member:

denotes a discrete cash outflow at any future time, .

Then, satisfaction of the goal necessitates a liquid portfolio amount greater than the required cash flow in each period. Conversely, if funds are insufficient to cover the cash outflow in any period, then the portfolio has failed to achieve the goal. Therefore, the probability of success can be represented using the following piecewise function:

signifies the probability of achieving a goal and is the liquid portfolio amount at a future time.

is stochastic; consequently, we simulate it via Monte Carlo and estimate based on the results as follows:

indexes the Monte Carlo trials and is the total number of trials.

**Appendix B: Forecasting Portfolio Outcomes Based on Scenarios**

The forecast consist of a strictly monotonically increasing sequence of five constant scenarios:

For convenience, these scenarios are labeled as follows:

The Cumulative Probability Density Function for the forecast is assumed to conform to the following constraint:

In words, and are constants that determine the area under the left and right tails of the Probability Density Function.

In addition, the Probability Density Function for the forecast is assumed to be piecewise linear with finite bounds:

Since and are triangles, and are calculated based on the scenarios and area constants:

The probability between and represents the area of the distribution not consisting of and . A vertical line drawn from to splits this area into two trapezoids. As a result, is determined by the aforementioned areas, heights, and scenarios:

The two trapezoidal areas can also be named and calculated:

The Piecewise Probability Density Function is fully specified based on the above results:

**Appendix C: Calculating Moments of the Forecast**

Through a property of expectations, the expectation of a piecewise linear function is the sum of its parts:

Thus, the first four moments of the distribution may be calculated analytically using the following definitions of the moments:

**Appendix D: Bootstrapping the Base Forecast**

Suppose a sample of historical returns daily returns, , has been recorded for an investment or its proxy (e.g. an index)

The distribution of historical returns of a longer time period, such as a year, can be approximated through statistical bootstrapping of the daily returns. In other words, a subsample of the daily historical returns can be chosen with replacement and then compounded to mimic annual returns.

is the estimated annual return based on bootstrapping, is the index of the randomly sampled first daily return, is the index of the randomly sampled second daily return, and represents the total number of samples chosen. If there are approximately 252 trading days in a year, then would equal 252.

If the above bootstrapping is repeated, then the result is a sequence of estimated annual returns .

Descriptive statistics such as the central moments and percentiles can be obtained using standard estimators and formulas. For example, the Minimum annual return for the base forecast is simply the minimum observed value of .

**Appendix E: Simulating Investment Performance based on the Forecasts**

The Piecewise Cumulative Density Function can also be determined:

Moreover, the above CDFs are quadratic. Consequently, the Inverse CDFs are calculated via the quadratic formula. For example, the Inverse CDF of the first piecewise segment can be calculated as follows:

Consequently, for any randomly generated , it is possible to map to the corresponding . The remaining Inverse CDFs are calculated similarly.