Modeling A Systematic Withdrawal Plan: A Stochastic Algorithm to Estimate Initial Fund Requirement

Shreeyans Dhamane

1. Introduction

To oppose inflation's depreciating effect on retirement savings, individuals seek to invest their savings to ensure appreciation over time. A systematic withdrawal plan refers to monthly withdrawals from an appreciating fund, especially to provide retirement income [1]. After a certain duration, the increasing withdrawals due to inflation will surpass the income earned by the appreciation rate of the retirement fund. This will cause the fund's value to reach zero. Hence, the aim of this study is to develop an algorithm to find the initial amount required to sustain a systematic withdrawal plan invested in the S&P 500 for a target duration, given an initial monthly withdrawal amount. Since the S&P 500 is widely diversified across several industries, individuals seek to invest their retirement funds in S&P 500 index funds to minimize risk caused by random chance [2]. Nevertheless, the S&P 500 displays randomness. To model systematic withdrawal plans accurately, it is vital to investigate if this randomness in the monthly growth rate of the S&P 500, as shown in Figure 1, is stochastic. Moreover, systematic withdrawal plans are more vulnerable to randomness in the S&P 500 than regular investments, as low growth rates in the initial months can cause the fund to run out quicker than expected since money is withdrawn from the fund periodically. Hence, to account for the higher risk posed by systematic withdrawal plans, a large safety margin must be determined by comparing a deterministic function modeling the fund with the stochastic simulations modeling the fund. Inflation rate fluctuations in the USA are trivial compared to the S&P 500 fluctuations; hence, the monthly inflation rate is assumed to be constant at 1.0021 (Average monthly inflation rate from 2010 to 2024) [3]. Section 2 involves deriving a deterministic function to model the systematic withdrawal plan, which uses the mean monthly growth rate of the S&P 500 from 2010 to 2024 as a constant. This function is used in the stochastic algorithm in Section 3. Section 3 involves exploratory data analysis to create a log-normal distribution of the fund's longevity. The stochastic algorithm in this study is a handy tool for estimating the initial fund requirement with 90% certainty based on the target duration and withdrawal amounts.

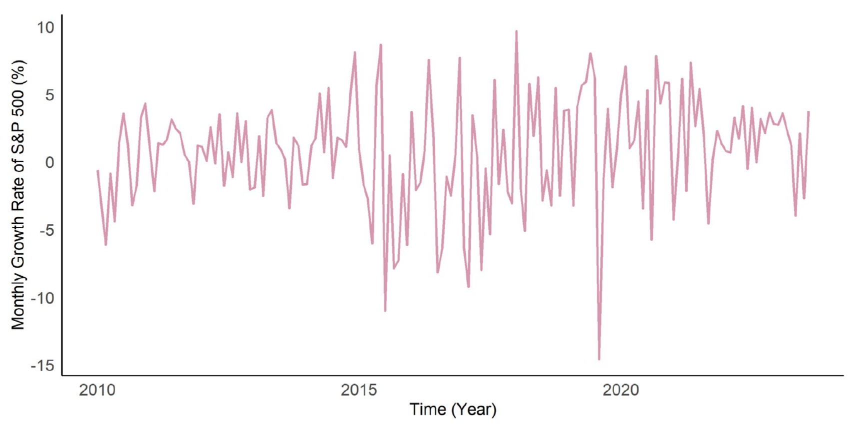

Figure 1. Monthly growth rate of the S&P 500 (2010-2024).

2. Deterministic Function to Model a Systematic Withdrawal Plan

2.1. Variables

We derive a deterministic function to model a systematic withdrawal plan. Table 1 outlines the variables used in this function.

Table 1. Definitions of the variables.

Variable	Definition
α	The initial amount in the fund
r_g	The monthly growth rate of the S&P 500
β	The initial monthly withdrawal amount
r_i	The monthly inflation rate in the United States

2.2. Deriving Expressions for Months 1 and 2

Procedure for Month 1:

Step 1: We multiply the initial amount in the fund α by $(1+r_g/100)$ to calculate the appreciated amount at Month 1's end:

$$\alpha\left(1+\frac{r_g}{100}\right)$$

Step 2: We multiply the initial monthly withdrawal amount β by $(1+r_i/100)$ to calculate the amount to be withdrawn at Month 1's end:

$$\beta\left(1+\frac{r_i}{100}\right)$$

Step 3: We subtract the amount to be withdrawn at Month 1's end from the appreciated amount at Month 1's end:

$$f(1)=\alpha\left(1+\frac{r_g}{100}\right)-\beta\left(1+\frac{r_i}{100}\right) \tag{1}$$

Step 4: We set the amount at Month 1's end to the amount at the start of Month 2.

Procedure for Month 2:

Step 1: We multiply the amount at the start of Month 2 by $(1+r_g/100)$ to calculate the appreciated amount at Month 2's end:

$$\left(\alpha\left(1+\frac{r_g}{100}\right)-\beta\left(1+\frac{r_i}{100}\right)\right)\left(1+\frac{r_g}{100}\right)$$

Step 2: We multiply the amount to be withdrawn at Month 1's end by $(1+r_i/100)$ to calculate the amount to be withdrawn at Month 2's end:

$$\beta\left(1+\frac{r_i}{100}\right)^2$$

Step 3: We subtract the amount to be withdrawn at Month 2's end from the appreciated amount at Month 2's end, which gives

$$\begin{aligned} f(2)&=\left(\alpha\left(1+\frac{r_g}{100}\right)-\beta\left(1+\frac{r_i}{100}\right)\right)\left(1+\frac{r_g}{100}\right)-\beta\left(1+\frac{r_i}{100}\right)^2 \\ &=\alpha\left(1+\frac{r_g}{100}\right)^2-\left(1+\frac{r_g}{100}\right)\beta\left(1+\frac{r_i}{100}\right)-\beta\left(1+\frac{r_i}{100}\right)^2 \end{aligned} \tag{2}$$

Step 4: We set the amount at Month 2's end to the amount at the start of Month 3. At this stage, two simplifications avoid excessive complexity in the model equation.

Simplification 1: A variable m_g is set to the value $\left(1+\frac{r_g}{100}\right)$.

Simplification 2: A variable m_i is set to the value $\left(1+\frac{r_i}{100}\right)$.

Using the simplifications, the expression for the fund's value at Month 1's end is:

$$f(1)=\alpha m_g-\beta m_i \tag{3}$$

Using the simplifications, the expression for the fund's value at Month 2's end is:

$$f(2)=\alpha m_g^2-m_g\beta m_i-\beta m_i^2 \tag{4}$$

The procedure used for Months 1 and 2 is repeated for Months 3, 4, 5, and so on until the fund reaches zero.

2.3. Deriving Expressions for Every Month

Table 2 illustrates the expression for the fund's value at every month's end after withdrawal. The first five months have been included in the table for representation purposes.

Table 2. The amount in the fund at the month's end after withdrawal.

Month	The amount in the fund at the month's end after withdrawal
1	$\alpha m_g-\beta m_i$
2	$\alpha m_g^2-m_g\beta m_i-\beta m_i^2$
3	$\alpha m_g^3-m_g^2\beta m_i-m_g\beta m_i^2-\beta m_i^3$
4	$\alpha m_g^4-m_g^3\beta m_i-m_g^2\beta m_i^2-m_g\beta m_i^3-\beta m_i^4$
5	$\alpha m_g^5-m_g^4\beta m_i-m_g^3\beta m_i^2-m_g^2\beta m_i^3-m_g\beta m_i^4-\beta m_i^5$

We factor out β from every term except the first term.

For example, for Month 3, we have

$$\begin{aligned} f(3)&=\alpha m_g^3-m_g^2\beta m_i-m_g\beta m_i^2-\beta m_i^3 \\ &=\alpha m_g^3-\beta\left(m_g^2 m_i+m_g m_i^2+m_i^3\right) \end{aligned} \tag{5}$$

We highlight the exponents of m_g and m_i in every term. Table 3 displays the amount in the fund at the month's end with highlighted exponents.

Table 3. The amount in the fund at the month's end with highlighted exponents.

Month	The amount in the fund at the month's end after withdrawal
1	$\alpha m_g^1-\beta\left(m_g^0 m_i^1\right)$
2	$\alpha m_g^2-\beta\left(m_g^1 m_i^1+m_g^0 m_i^2\right)$
3	$\alpha m_g^3-\beta\left(m_g^2 m_i^1+m_g^1 m_i^2+m_g^0 m_i^3\right)$
4	$\alpha m_g^4-\beta\left(m_g^3 m_i+m_g^2 m_i^2+m_g^1 m_i^3+m_g^0 m_i^4\right)$
5	$\alpha m_g^5-\beta\left(m_g^4 m_i+m_g^3 m_i^2+m_g^2 m_i^3+m_g^1 m_i^4+m_g^0 m_i^5\right)$

2.4. Splitting Expressions for Every Month

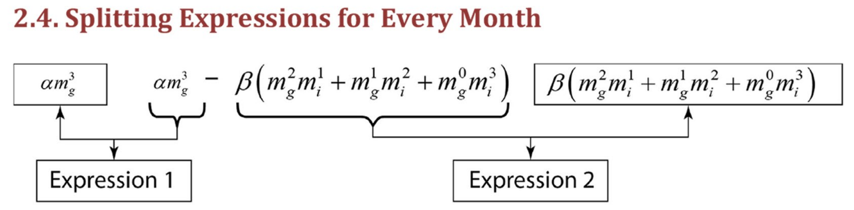

Figure 2. Splitting of the expression (Example for Month 3).

For every month, the expression can be written as a subtraction of two distinct expressions: Expression 1 and Expression 2 (Figure 2).

2.5. Modeling Expression 1

For every month, Table 4 illustrates Expression 1 and the exponent of the variable m_g in Expression 1.

Table 4. Expression 1 and the exponent of m_g in Expression 1.

Month	Expression 1	Exponent of variable m_g
1	am_g^1	1
2	am_g^2	2
3	am_g^3	3
4	am_g^4	4
5	am_g^5	5

Let the variable x represent the month number. For example, the value of x for Month 3 is 3.

Then, Expression 1 can be modeled by the function:

$$f(x)=\alpha m_g^x \tag{6}$$

2.6. Modeling Expression 2

For every month, it is evident that the expression that the variable β multiplies is a series (Figure 3).

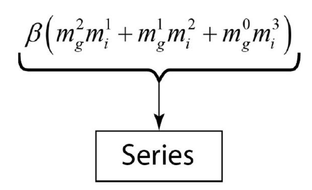

Figure 3. The expression that β multiplies is a series (Example for Month 3).

Step 1: We isolate the series from the variable β.

Step 2: We isolate every term of the series.

Table 5. Isolation of every term of the series.

Month	Terms of the series				
	1st Term	2nd Term	3rd Term	4th Term	5th term
1	$m_g^0 m_i^1$	-	-	-	-
2	$m_g^1 m_i^1$	$m_g^0 m_i^2$	-	-	-
3	$m_g^2 m_i^1$	$m_g^1 m_i^2$	$m_g^0 m_i^3$	-	-
4	$m_g^3 m_i^1$	$m_g^2 m_i^2$	$m_g^1 m_i^3$	$m_g^0 m_i^4$	-
5	$m_g^4 m_i^1$	$m_g^3 m_i^2$	$m_g^2 m_i^3$	$m_g^1 m_i^4$	$m_g^0 m_i^5$

Step 3: We isolate the exponent of the variable m_g and the exponent of the variable m_i from every term of the series (Table 5).

Table 6. Isolation of the exponents of m_g and m_i from every term of the series.

Month	Term	Exponent of m_g (j)	Exponent of m_i (k)	Exponent of m_g + Exponent of m_i ($j + k$)
1	1st Term	0	1	1
2	1st Term	1	1	2
	2nd Term	0	2	2
3	1st Term	2	1	3
	2nd Term	1	2	3
	3rd Term	0	3	3
4	1st Term	3	1	4
	2nd term	2	2	4
	3rd Term	1	3	4
	4th Term	0	4	4
5	1st Term	4	1	5
	2nd Term	3	2	5
	3rd Term	2	3	5
	4th Term	1	4	5
	5th Term	0	5	5

Step 4: By observing Table 6, we deduce that the month number is equal to the number of terms in the series corresponding to that month. Then, the variable x (defined in Section 2.5) represents both the month number and the number of terms in the series corresponding to that month.

Step 5: Let j denote the exponent of the variable m_g, and let k denote the exponent of the variable m_i

Step 6: Therefore, by observing the trends in the values of j and k in Table 7, we formulate a generalized expression for representing the series for any arbitrary month x:

$$\sum_{k=1}^{x} m_g^j m_i^k$$

Step 7: By observing the rightmost column of Table 6, we can deduce that the sum of j and k for any arbitrary month x is equal to the month number x.

Then, we have:

$$\begin{aligned} j+k&=x \\ j&=x-k \end{aligned} \tag{7}$$

We substitute j with $x-k$ in the generalized expression for the series:

$$\sum_{k=1}^{x} m_g^{x-k} m_i^k$$

Step 8: We combine the generalized expression for the series with the variable β, which was isolated in Step 1, to obtain the function modelling Expression 2:

$$f(x)=\beta\left(\sum_{k=1}^{x} m_g^{x-k} m_i^k\right) \tag{8}$$

2.7. Formulating the Function

As shown in Figure 2, the main expression is formed by subtracting Expression 2 from Expression 1.

Therefore, the function modelling the main expression:

$$f(x)=am_g^x-\beta\left(\sum_{k=1}^{x} m_g^{x-k} m_i^k\right) \tag{9}$$

We replace our simplifications m_g and m_i with their original expansions to obtain:

$$f(x)=\alpha\left(1+\frac{r_g}{100}\right)^x-\beta\left(\sum_{k=1}^{x}\left(1+\frac{r_g}{100}\right)^{x-k}\left(1+\frac{r_i}{100}\right)^k\right) \tag{10}$$

The function above is a function of the month number x, and $f(x)$ represents the amount in the fund at the end of x months. The function models the fund of a systematic withdrawal plan, for which the variables α, β, r_g, and r_i can be inputted.

3. Stochastic Analysis and Simulation of Fund Longevity

3.1. Exploratory Data Analysis

The function that models the fund assumes that r_g is constant. However, r_g, which denotes the monthly growth rate of the S&P 500, is bound to vary. Therefore, we must account for the uncertainty in r_g in the model. A dataset of 168 r_g values was retrieved [4]. Each r_g value corresponds to its respective month from January 2010 to January 2024. To check if the randomness in r_g values follows a predictable pattern, exploratory data analysis is conducted by generating a histogram of r_g values (Figure 4).

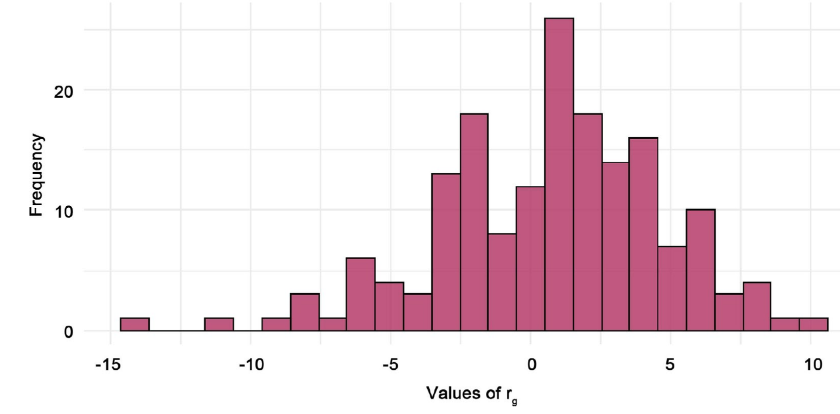

Figure 4. Histogram of r_g values.

Observing the shape of the histogram, we can intuitively hypothesize that the variable r_g is likely to be normally distributed. To validate our hypothesis, we shall check the normality of r_g using a one-sample Kolmogorov–Smirnov test [5].

3.2. Kolmogorov–Smirnov Test

Hypotheses for the one-sample Kolmogorov–Smirnov test on r_g:

H: r_g is a random variable drawn from a normal distribution with μ equal to the mean of r_g and σ equal to the standard deviation of r_g.

H_A: r_g is not a random variable drawn from a normal distribution with μ equal to the mean of r_g and σ equal to the standard deviation of r_g.

Table 7. Kolmogorov-Smirnov test results.

Variable	Value
KS Statistic	0.0764
p-value	0.2582
μ	0.6933 (% per month)
σ	4.0266 (% per month)

The Kolmogorov–Smirnov test outputs a KS Statistic of 0.0764 and a p-value of 0.2582. Since the p-value of 0.2582 is more than the alpha value of 0.05, the null hypothesis is accepted. This comparison proves that r_g is random variable normally distributed with μ as 0.6933 (% per month) and σ as 4.0266 (% per month).

3.3. Stochastic Simulations

Since we know that r_g is a normally distributed random variable, we can run stochastic simulations for the function that models the fund. Each time we run a stochastic simulation; we note the number of months it takes for the function to fall below zero. In other words, we note how long the fund of a systematic withdrawal plan lasts. Let t be a new variable, where t refers to the number of months it takes for the fund to fall below zero. Since the function that models the fund depends on r_g, the uncertainty in r_g carries over to t. By running sufficient stochastic simulations, we can generate a histogram of t values to check if the randomness in t values follows a predictable pattern. We set realistic example values for α as $500000 and β as $5000 for the stochastic simulations. r_g is sampled from a normal distribution with parameters (0.6933, 4.0266), and r_i is 1.0021(Figure 5).

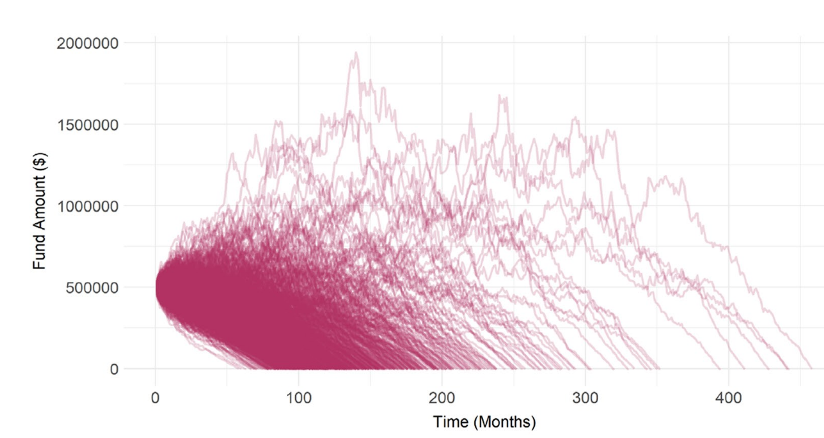

Figure 5. 10000 simulations: $\alpha = 500000$, $\beta = 5000$, r_g = normal (0.6933, 4.0266), $r_i = 1.0021$.

Pseudocode for plotting a histogram of t values by running stochastic simulations:

```
loop i from 1 to 10000
    f(x) = α
  loop x while f(x) > 0
    rg = np.random.normal(mu, sigma)
    f(x) = (f(x)*(1 + (rg/100))) – (β*(1 + (ri/100))^x)
    x++
  end loop
  t_values[i] = x
end loop
plot histogram of t_values
```

3.4. Outliers

Figure 6 suggests that outliers exist when t equals 10,000 months. These outliers are halted at 10,000 months due to the technical limitations of the R language. However, these outliers signify extreme cases when t approaches infinity. These cases happen when stochastic variations result in unrealistic consistently high r_g values, which allow the fund to last infinitely. However, these cases are significantly rare, with a 0.58% (58 in 10,000 simulations) probability. Therefore, we remove the outliers to improve robustness, and we plot a new histogram.

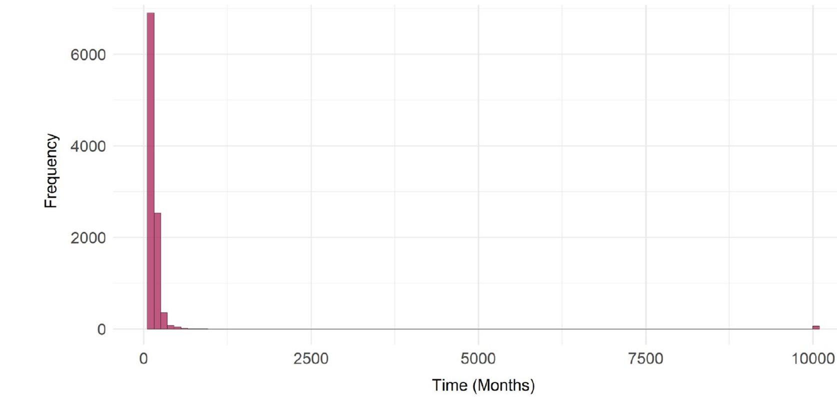

Figure 6. Histogram of t values with outliers.

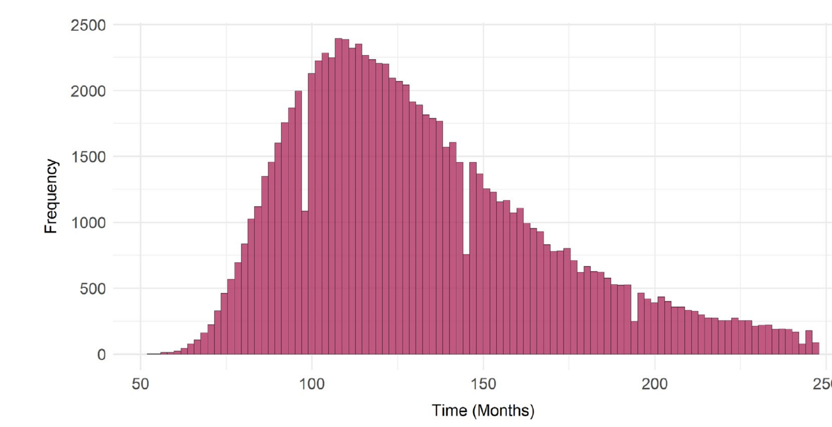

Figure 7. Histogram of t values without outliers.

The shape of the histogram of t values in Figure 7 suggests that the variable t is likely to be log-normally distributed. To validate our hypothesis, we shall log-transform [6] the t values and check the normality of the log-transformed t values using a one-sample Kolmogorov–Smirnov test.

Table 8. Kolmogorov-Smirnov test results.

Variable	Value
KS Statistic	0.0488
p-value	0.1318
μ_t	4.8466
σ_t	0.2721

The Kolmogorov–Smirnov test outputs a KS Statistic of 0.0488 and a p-value of 0.1318. Since the p-value is more than the alpha value of 0.05, the null hypothesis is accepted (Table 8). This comparison proves that t is a random variable drawn from a log-normal distribution with μ_t equal to 4.8466 and σ_t equal to 0.2721. Therefore, a continuous probability distribution function for t values, based on the log-normal distribution [7] is given by:

$$P(x)=\frac{1}{2}\left(1+\operatorname{erf}\left(\frac{\ln(x)-4.8466}{0.2721\sqrt{2}}\right)\right) \tag{11}$$

Equation (11) applies specifically to the example given in this study. However, the log-normal nature of the distribution of t values holds for all values of α and β, as the shape of the distribution depends on r_g, and r_g follows a fixed normal distribution.

3.5. Percentiles

Table 9. Percentiles of t values.

Percentile	t Value (Months)[a]
10th	90
20th	101
30th	110
40th	119
50th	127
60th	136
70th	147
80th	160
90th	180

a. t values are rounded to the nearest integer.

Table 9 displays the percentiles of t values according to the log-normal distribution given by Equation (11). The deterministic model from Section 2, which does not account for randomness, predicts a t value of 137 months for $\alpha = 500{,}000$, $\beta = 5000$, $r_g = 0.6933$, and $r_i = 1.0021$. r_g was set to 0.6933, which is the mean and best estimate of r_g, as the model requires a constant value. This prediction is validated by our stochastic analysis, as 137 is close to 127, which is the median of the log-normal distribution of t values. However, the deterministic model and the median are not a reasonable estimate of the fund's longevity; they approximately provide a 50% safety margin, as 50% of the time, we will get a t value of less than 127 months, solely due to random chance. We must have a 90% safety margin, chosen in several equity funds, to have a high level of confidence in the fund's longevity.

3.6. Stochastic Algorithm

Our goal is to develop an algorithm to output α, the initial fund requirement, with a 90% safety margin, based on inputs of β and t_{Target} Let t_{Target} be a new variable, which represents the number of months we require our fund to last.

Step 1: Given inputs β and t_{Target}, we use the deterministic model from Section 2 to find the 50% safety margin value for α.

When we set Equation (10) from Section 2 to equal zero, the x in the equation represents the number of months the fund of the systematic withdrawal plan lasts.

$$\alpha\left(1+\frac{r_g}{100}\right)^x-\beta\left(\sum_{k=1}^{x}\left(1+\frac{r_g}{100}\right)^{x-k}\left(1+\frac{r_i}{100}\right)^k\right)=0 \tag{12}$$

We rearrange Equation (13) to obtain:

$$\alpha\left(1+\frac{r_g}{100}\right)^x=\beta\left(\sum_{k=1}^{x}\left(1+\frac{r_g}{100}\right)^{x-k}\left(1+\frac{r_i}{100}\right)^k\right)$$

$$\alpha=\frac{\beta\left(\sum_{k=1}^{x}\left(1+\frac{r_g}{100}\right)^{x-k}\left(1+\frac{r_i}{100}\right)^k\right)}{\left(1+\frac{r_g}{100}\right)^x}$$

$$\alpha=\frac{\beta\left(\sum_{k=1}^{t_{target}}\left(1+\frac{r_g}{100}\right)^{t_{target}-k}\left(1+\frac{r_i}{100}\right)^k\right)}{\left(1+\frac{r_g}{100}\right)^{t_{target}}} \tag{13}$$

We plug the values of β, $r_g = 0.6933$, $r_i = 1.0021$, and t_{Target} into Equation (13) to calculate the 50% safety value of α.

Step 2: We run stochastic simulations to find the log-normal distribution of t values for the 50% safety value of α using the pseudocode from Section 3.3.

Step 3: We identify the 90th percentile t value from the log-normal distribution of t values, and we plug it into Equation (13) to find the 90% safety value of α.

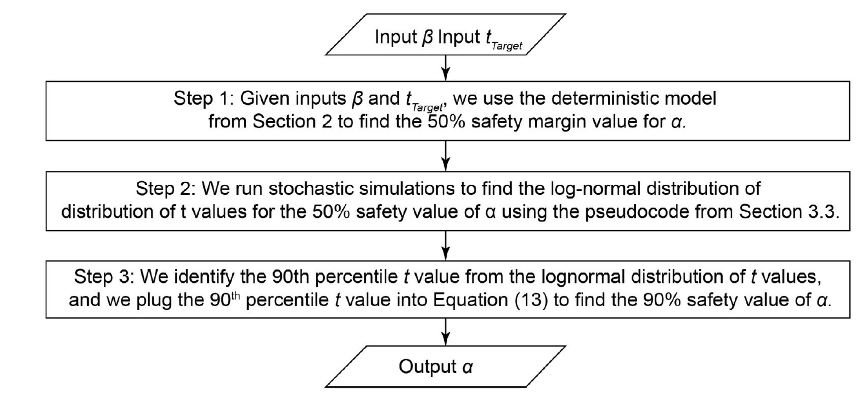

Figure 8. Flowchart for the stochastic algorithm.

For example, we set the same example value from Section 3.3 of β as $5000, and we set the target t value as 127 months from Section 3.5. Running the stochastic algorithm from Figure 8 on these parameters returns α as $629,065. Compared to the lower initial α value of $500,000 from Section 3.3, which had a 50% safety margin, the higher α value of $629,065 suggested by our algorithm gives a 90% safety margin.

4. Conclusion

This study identified that the monthly growth rate of the S&P 500 is normally distributed, with a mean of 0.6933% per month and a standard deviation of 4.0266% per month. The algorithm performs efficiently for other stocks indexes like BSE SENSEX and FTSE 100, provided their monthly growth rate is normally distributed. Stochastic simulations using this normal distribution led to the observation of a log-normal distribution of the fund's longevity after the removal of outliers. Based on the distribution of the fund's longevity, this study proposes a stochastic algorithm that output the initial amount required to sustain a systematic withdrawal plan for a target duration with a 90% safety margin, given an initial monthly withdrawal amount. Applying the algorithm to countries with unstable inflation and stock fluctuations [8] may need larger safety margins of 95% to account for increased randomness. The algorithm outperforms conventional systematic withdrawal plan calculators since it is personalized by inputting the mean and standard deviation for any stock. A benefit of the algorithm is the removal of unrealistic favorable outliers. Therefore, the predictions of the initial fund requirements are not skewed by extreme data points [9]. Another advantage of the algorithm is its 90% safety margin, which ensures that retirement planners have a high level of confidence that the fund of their systematic withdrawal plan will last for their target durations.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Shah, T. and Baser, N. (2022) Global Mutual Fund Market: The Turn of the Month Effect and Investment Strategy. *Journal of Asset Management*, **23**, 466-476. https://doi.org/10.1057/s41260-022-00282-0

[2] Sodini, P. and Viceira, L.M. (2020) The Value of Diversification: Diversification is the Only Free Lunch in Finance. Harvard University. https://scholar.harvard.edu/files/lviceira/files/ap7_annual_report-ps_and_lv-2020-01-29.pdf

[3] Macrotrends (2024) U.S. Inflation Rate 1960-2024. Macrotrends. https://www.macrotrends.net/global-metrics/countries/USA/united-states/inflation-rate-cpi

[4] Investing.com (2024) S&P 500 Index Historical Data. https://www.investing.com/indices/us-spx-500-historical-data

[5] Bristol University (2024) SPSS—Exploring Normality (Practical). https://www.bristol.ac.uk/cmm/media/research/ba-teaching-