

An Economic Analysis of Optimal Investment Strategies for Accumulating Housing Down Payments

Business Analytics MS Capstone Project

Frank Paul Longo II

July 25, 2024

Abstract

This research provides a comprehensive analysis of optimal investment strategies for first-time homebuyers aiming to accumulate down payments over 5, 7.5, and 10-year horizons. By leveraging Modern Portfolio Theory, the Capital Asset Pricing Model, and Monte Carlo simulations, the study offers actionable insights into constructing age-specific investment portfolios.

The results indicate that tailored investment strategies significantly enhance the ability to save for a down payment, reducing the time required to reach homeownership goals. The findings also highlight the importance of considering risk-adjusted returns and diversification in portfolio construction.

Future research could extend this analysis to consider other demographic factors such as income levels and regional housing market conditions. Additionally, the integration of alternative investment vehicles such as real estate investment trusts (REITs) and cryptocurrencies could provide further insights into optimizing investment strategies for down payments.

Contents

1	Introduction	5
1.1	Research Question & Objective	5
1.2	Motivation	5
2	Literature Review	6
2.1	Foundational Theories	6
2.1.1	Markowitz's Modern Portfolio Theory	6
2.1.2	Sharpe's Capital Asset Pricing Model	6
2.2	Monte Carlo Simulations	6
2.2.1	Monte Carlo in Portfolio Management	6
2.2.2	Monte Carlo in Real Estate Investments	7
2.3	Gaps in the Literature	7
3	Data	7
3.1	Date Range	7
3.2	Data Fields	7
4	Theoretical Models	9
4.1	Capital Asset Pricing Model (CAPM)	9
4.1.1	Beta Calculation	9
4.1.2	Security Market Line (SML)	9
4.1.3	Theoretical Implications	9
4.2	Modern Portfolio Theory (MPT)	10
4.2.1	Efficient Frontier and Optimal Portfolio	10
4.3	Monte Carlo Simulation	10
5	Empirical Specification	12
5.1	Model Implementation	12
5.1.1	Data Collection and Preparation	12
5.1.2	CAPM and Sharpe Ratio Calculation	12
5.1.3	Composite Score Calculation	13
5.1.4	Modern Portfolio Theory (MPT) Application	13
5.1.5	Gathering Hindsight Data	13

5.1.6	Comparison with Hindsight Data	13
5.1.7	Monte Carlo Simulation for Future Forecasting	13
6	Results	15
6.1	Initial Data Analysis	15
6.1.1	Distribution of Securities by Type	15
6.1.2	Cumulative Returns by Security Type	15
6.2	Optimal Portfolios	15
6.2.1	Top Assets by Composite Score	15
6.2.2	Optimal Portfolios Composition	18
6.3	Comparison with Hindsight Data	18
6.3.1	Cumulative Returns Comparison	18
6.3.2	Cumulative Returns Summary	20
6.4	Monte Carlo Simulation for Future Forecasting	21
6.4.1	Distribution of Final Portfolio Values	21
6.4.2	Cumulative Returns Over Time	21
6.5	Summary Statistics	24
7	Conclusions	25
A	Data Appendix	26
A.1	Summary of Data Sources	26
A.2	Data Cleaning and Processing Details	26
A.3	Python Scripts for Data Processing	26
A.4	Variable Definitions	26
B	Acknowledgments	28

1 Introduction

The escalating housing costs in contemporary real estate markets have created significant barriers for first-time homebuyers. This demographic often faces the daunting task of accumulating substantial down payments amidst economic volatility and uncertain income trajectories. This research addresses this critical issue by developing and evaluating optimal investment strategies tailored to help diverse age groups achieve their homeownership goals within 5, 7.5, and 10-year horizons.

1.1 Research Question & Objective

The central research question guiding this investigation is: What are the most effective investment strategies for different age groups to accumulate housing down payments over periods of 5, 7.5, and 10 years? The primary objective of this study is to identify, analyze, and optimize investment strategies that can effectively assist first-time homebuyers in saving for their down payments. By leveraging advanced financial theories and empirical methodologies, this research aims to provide actionable insights that balance risk and return, offering practical solutions for prospective homeowners (Markowitz, 1952; Sharpe, 1966; Boyle, 1977).

1.2 Motivation

The motivation for this research stems from the pressing need to address the challenges posed by rising housing costs and economic instability. As homeownership becomes increasingly out of reach for many, particularly younger individuals, it is imperative to develop strategies that can mitigate these barriers. By providing evidence-based investment strategies, this study aims to empower individuals with the tools needed to navigate the complexities of financial planning for homeownership (???)

2 Literature Review

The literature on optimal investment strategies and their applications in real estate and housing markets is extensive. Seminal works by Markowitz (1952) on Modern Portfolio Theory (MPT) and ? on the Capital Asset Pricing Model (CAPM) provide the foundational theories. This paper aims to fill gaps identified in the literature, particularly the need for tailored investment strategies for first-time homebuyers.

2.1 Foundational Theories

2.1.1 Markowitz's Modern Portfolio Theory

Markowitz's Modern Portfolio Theory (Markowitz, 1952) introduced the concept of portfolio optimization, emphasizing the trade-off between risk and return. The theory suggests that investors can achieve optimal portfolios by diversifying their investments across different assets, thereby reducing risk without sacrificing expected returns. Markowitz's efficient frontier illustrates the set of optimal portfolios that offer the highest expected return for a given level of risk.

2.1.2 Sharpe's Capital Asset Pricing Model

Sharpe's Capital Asset Pricing Model (?) extends the notion of risk by introducing systematic and unsystematic risk. The model posits that the expected return on an asset is a function of its sensitivity to market movements (beta), the risk-free rate, and the market risk premium. Sharpe's work was pivotal in differentiating between diversifiable (unsystematic) risk and non-diversifiable (systematic) risk, providing a framework for understanding how different types of risk impact asset pricing and returns.

2.2 Monte Carlo Simulations

Monte Carlo simulations have become a critical tool in financial modeling, providing a method for assessing the impact of risk and uncertainty in investment strategies. Boyle (Boyle, 1977) introduced the Monte Carlo approach to option pricing, and since then, its applications have expanded across various fields of finance.

2.2.1 Monte Carlo in Portfolio Management

Monte Carlo simulations are used to model the behavior of investment portfolios under a wide range of possible future scenarios. This method involves generating a large number of random samples from the probability distributions of asset returns and analyzing the resulting portfolio performance. Studies by Glasserman (2004) and others have demonstrated how Monte Carlo simulations can provide insights into

the risk and return profiles of different investment strategies, helping investors to better understand the potential outcomes and make more informed decisions.

2.2.2 Monte Carlo in Real Estate Investments

In the context of real estate, Monte Carlo simulations have been used to evaluate the risk and return of property investments. For example, Brown and Matysiak (Brown and Matysiak, 2000) employed Monte Carlo methods to assess the uncertainty in real estate portfolio returns, while Kuhle and Alvaay (Kuhle and Alvaay, 2021) used simulations to analyze the impact of economic shocks on real estate prices. These applications illustrate the versatility of Monte Carlo simulations in addressing the unique risks associated with real estate investments.

2.3 Gaps in the Literature

Despite the extensive research on investment strategies, there remains a need for tailored approaches that address the specific financial goals and constraints of first-time homebuyers. This paper seeks to fill this gap by developing and evaluating optimal investment strategies that cater to the unique needs of aspiring homeowners across different age cohorts and investment horizons.

3 Data

The financial data used in this research is sourced from Yahoo Finance (Yahoo Finance, nd), which includes comprehensive information on roughly 150 securities consisting of stocks, mutual funds, and ETFs. This data source provides a rich dataset for analyzing the performance of different investment vehicles over time.

3.1 Date Range

The data covers the period from May 2011 to November 2014 for daily frequency and further hindsight data from May 2011 to July 2024. This timeframe allows for the analysis of recent trends and the performance of different asset classes in various market conditions.

3.2 Data Fields

The dataset includes the following fields:

- Open: Price at the beginning of the trading day.

- High: Peak price during the trading day.
- Low: Lowest price during the trading day.
- Close: Price at the end of the trading day.
- Adj Close: Closing price adjusted for dividends, stock splits, etc.
- Volume: Number of shares traded during a single trading day.
- Type: Security type (e.g., stock, ETF).

These fields provide a comprehensive view of the daily trading activities and price movements of different securities, essential for the analysis of investment performance and strategy development.

4 Theoretical Models

4.1 Capital Asset Pricing Model (CAPM)

The Capital Asset Pricing Model (CAPM) is employed to determine the expected return on an asset based on its systematic risk, as measured by beta (β_i). The CAPM formula is:

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f) \quad (1)$$

4.1.1 Beta Calculation

Beta (β_i) measures the volatility of an asset in relation to the market. It is calculated as:

$$\beta_i = \frac{Cov(R_i, R_m)}{\sigma_m^2} \quad (2)$$

4.1.2 Security Market Line (SML)

The Security Market Line (SML) is a graphical representation of the CAPM, showcasing the relationship between the expected return of an asset and its systematic risk, as measured by beta (β_i).

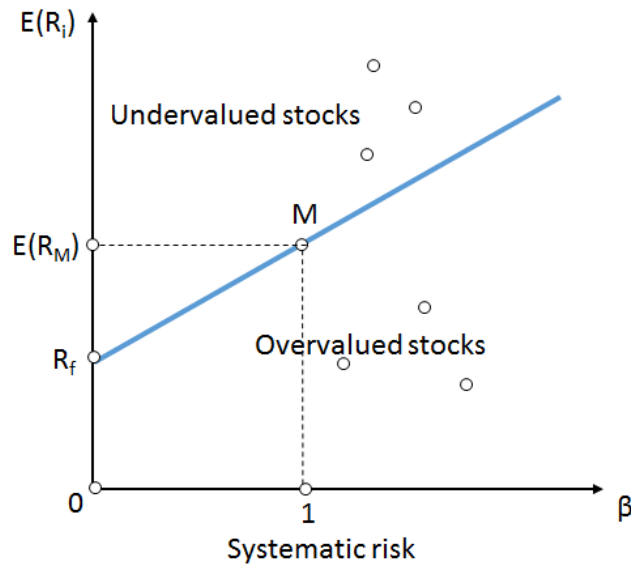


Figure 1: Security Market Line

4.1.3 Theoretical Implications

The SML conveys several important theoretical implications:

- All securities, when correctly priced, should lie on the SML.
- The slope of the SML is the market risk premium, $E(R_m) - R_f$, representing the additional return expected from holding a market portfolio instead of risk-free assets.
- The intercept of the SML is the risk-free rate, R_f , reflecting the return of a theoretically risk-free asset.

4.2 Modern Portfolio Theory (MPT)

Modern Portfolio Theory provides a robust framework for constructing an optimal portfolio that maximizes expected return for a given level of risk. The expected return $E(R_p)$ of a portfolio is the weighted sum of the expected returns of the individual assets:

$$E(R_p) = \sum_{i=1}^n w_i E(R_i) \quad (3)$$

4.2.1 Efficient Frontier and Optimal Portfolio

The efficient frontier is a concept from MPT that represents the set of optimal portfolios offering the highest expected return for a defined level of risk. The process of constructing the efficient frontier involves solving the following optimization problem:

$$\min \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij} \quad (4)$$

subject to:

$$\sum_{i=1}^n w_i = 1 \quad (5)$$

and

$$E(R_p) = \sum_{i=1}^n w_i E(R_i) \quad (6)$$

4.3 Monte Carlo Simulation

Monte Carlo simulations are utilized to model the uncertainty and variability in investment returns over time. The simulation process involves generating random returns based on historical data and iterating this process to build a distribution of potential outcomes. The value of an investment at time i is given by:

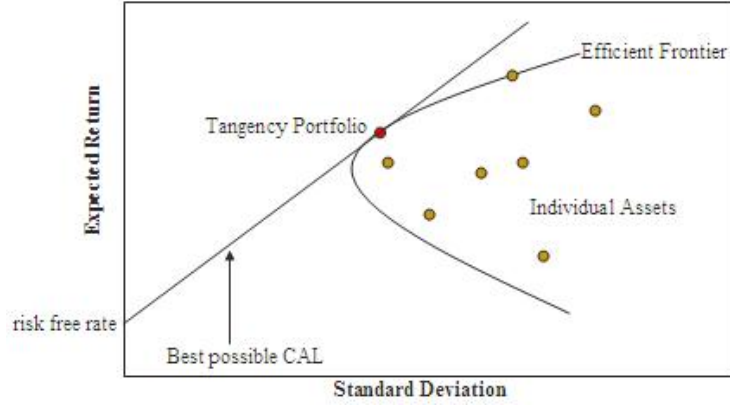


Figure 2: Efficient Frontier

$$X_i = X_{i-1} \times (1 + r_i) \quad (7)$$

where X_i is the investment value at time i and r_i is the return for period i . By running multiple simulations, we can estimate the expected value and variability of the investment portfolio, providing insights into the likelihood of achieving the desired down payment amount within the specified time horizon.

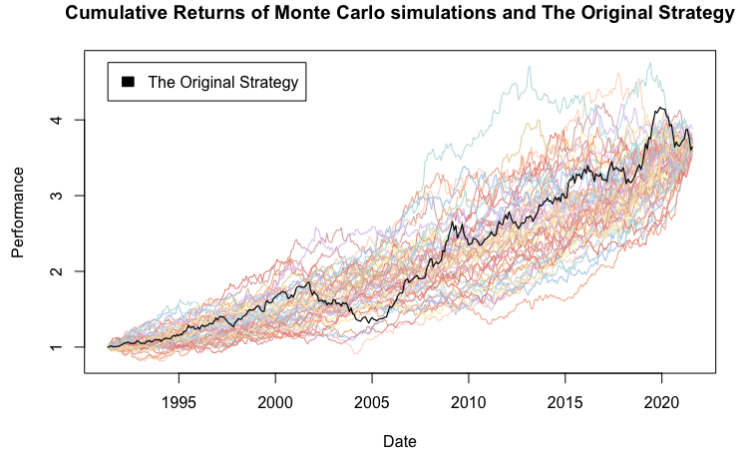


Figure 3: Investment Simulation Process for Monte Carlo Analysis

5 Empirical Specification

5.1 Model Implementation

The empirical analysis involves a systematic approach to evaluating and optimizing investment strategies using historical data. The process is outlined as follows:

5.1.1 Data Collection and Preparation

Data is collected from Yahoo Finance, covering a comprehensive range of securities, including stocks, ETFs, and mutual funds. The data spans from May 2011 to November 2014, with daily updates until July 2024. The data is cleaned and adjusted for corporate actions like stock splits and dividends using the `yfinance_data.py` script to ensure consistency (Yahoo Finance, nd).

5.1.2 CAPM and Sharpe Ratio Calculation

Each security's risk and return profile is assessed using the Capital Asset Pricing Model (CAPM) (?) and Sharpe Ratio (Sharpe, 1966). The process involves:

1. Calculating the average return of the market index (e.g., S&P 500) (?).
2. Determining the risk-free rate (e.g., 10-year U.S. Treasury bonds yield) (?).
3. Computing the beta (β) of each security:

$$\beta_i = \frac{\text{Cov}(R_i, R_m)}{\text{Var}(R_m)} \quad (8)$$

4. Estimating the expected return using the CAPM formula:

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f) \quad (9)$$

5. Calculating the Sharpe Ratio:

$$\text{Sharpe Ratio} = \frac{E(R_i) - R_f}{\sigma_i} \quad (10)$$

This analysis is performed using the `init_filtering.py` script.

5.1.3 Composite Score Calculation

Securities are ranked based on a composite score that integrates multiple metrics (?):

$$\text{Composite Score} = w_{\beta}\beta + w_{\text{Sharpe}}\text{Sharpe Ratio} + w_{\text{CAPM}}E(R_i) + w_{\text{Actual}}\text{Actual Returns} \quad (11)$$

Weights are assigned to each metric to reflect their importance in the ranking process.

5.1.4 Modern Portfolio Theory (MPT) Application

Using the ranked securities, portfolios are optimized for different investment horizons (5, 7.5, and 10 years) using Modern Portfolio Theory (MPT) (Markowitz, 1952). The optimization involves solving the following quadratic programming problem:

$$\text{Maximize} \quad \mathbf{w}^T \bar{\mathbf{r}} - \frac{\lambda}{2} \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} \quad (12)$$

$$\text{subject to} \quad \sum_i w_i = 1 \quad (13)$$

$$w_i \geq 0 \quad \forall i \quad (14)$$

where \mathbf{w} is the vector of asset weights, $\bar{\mathbf{r}}$ is the vector of expected returns, $\mathbf{\Sigma}$ is the covariance matrix of returns, and λ is the risk aversion parameter. This process is implemented using the `mpt.py` script.

5.1.5 Gathering Hindsight Data

Historical performance data from 2009 to the present is gathered and processed to evaluate the robustness of the optimized portfolios. This step uses the `hindsight_data.py` script to ensure that the data aligns by common dates and covers all relevant securities (?).

5.1.6 Comparison with Hindsight Data

The optimized portfolios are compared against actual historical performance using the hindsight data. The comparison includes calculating the cumulative returns of the portfolios and benchmarking them against the S&P 500 index (?). This step is performed using the `hindsight.py` script.

5.1.7 Monte Carlo Simulation for Future Forecasting

Monte Carlo simulations are conducted to forecast the performance of the optimized portfolios from today onward (Glasserman, 2004). The simulation process involves:

1. Defining initial investment amounts and annual contributions.
2. Generating random returns based on historical distributions.
3. Calculating portfolio values at each time step:

$$V_t = V_{t-1} \times (1 + R_t) + C \quad (15)$$

4. Running multiple iterations to build a probability distribution of outcomes.
5. Applying economic shocks to simulate real-world scenarios:

$$V_t = V_t \times (1 + \text{Shock Intensity}) \quad (16)$$

This comprehensive simulation is implemented using the `mcs.py` script, providing insights into the potential future performance of the portfolios (Brown and Matysiak, 2000; Kuhle and Alwayay, 2021).

6 Results

This section presents the results of the analysis, including the performance of the optimal portfolios over 5, 7.5, and 10-year horizons. The results are visualized using various figures and tables, organized to follow the empirical specification process.

6.1 Initial Data Analysis

6.1.1 Distribution of Securities by Type

Figure 4 shows the distribution of the securities by type.

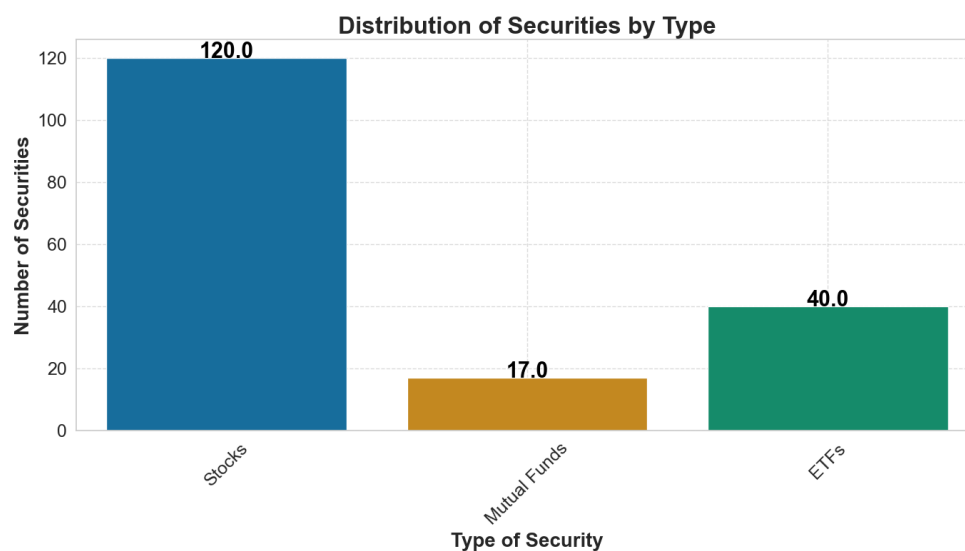


Figure 4: Distribution of Securities by Type

6.1.2 Cumulative Returns by Security Type

Figure 5 illustrates the cumulative returns over time by security type.

Interpretation: The majority of the securities analyzed are stocks, followed by ETFs and mutual funds. Stocks exhibit the highest cumulative return over time, indicating a higher potential for long-term growth compared to ETFs and mutual funds (Markowitz, 1952).

6.2 Optimal Portfolios

6.2.1 Top Assets by Composite Score

Figures 6, 7, and 8 show the top assets by composite score for 10, 7.5, and 5-year horizons, respectively.

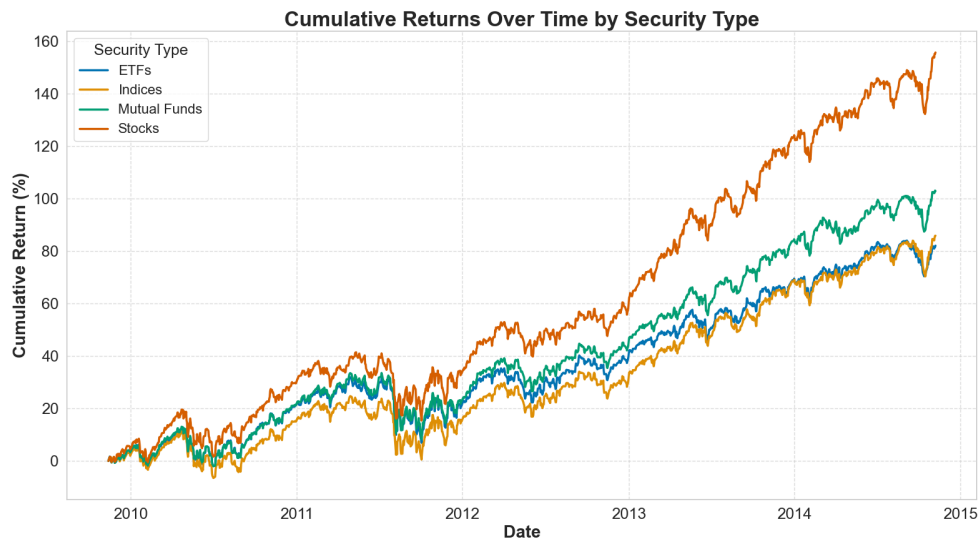


Figure 5: Cumulative Returns Over Time by Security Type

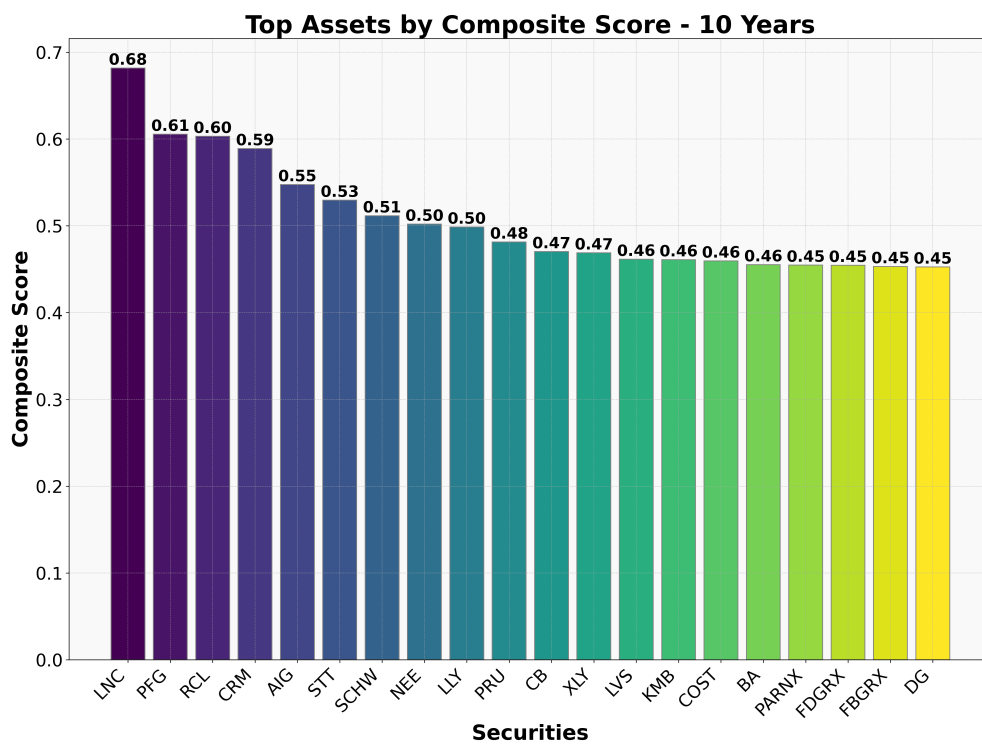


Figure 6: Top Assets by Composite Score (10 Years)

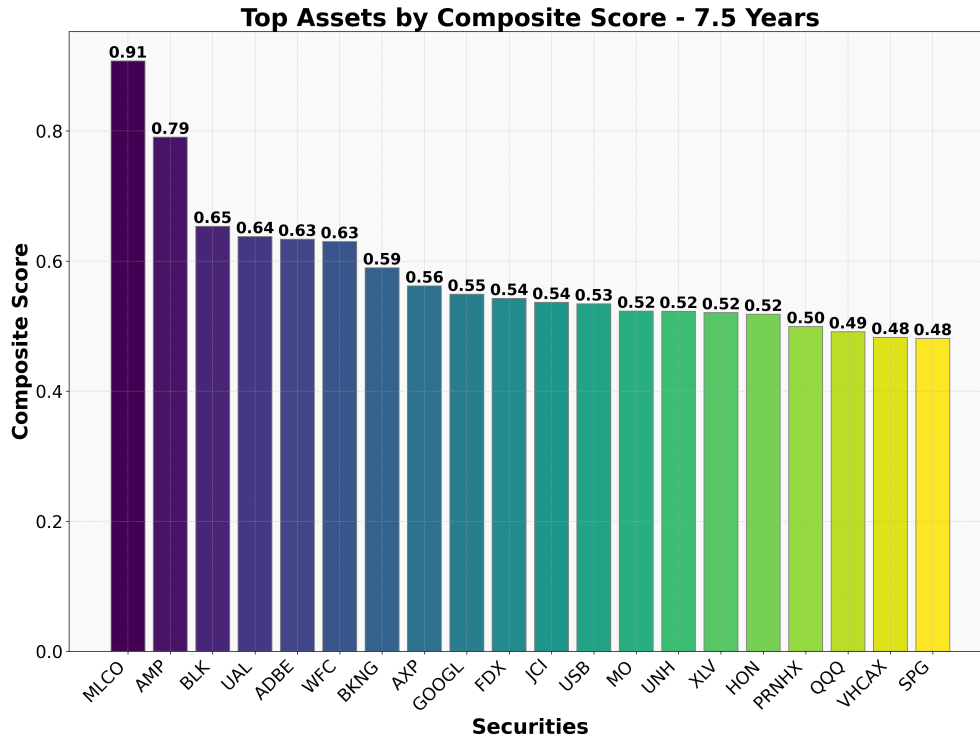


Figure 7: Top Assets by Composite Score (7.5 Years)

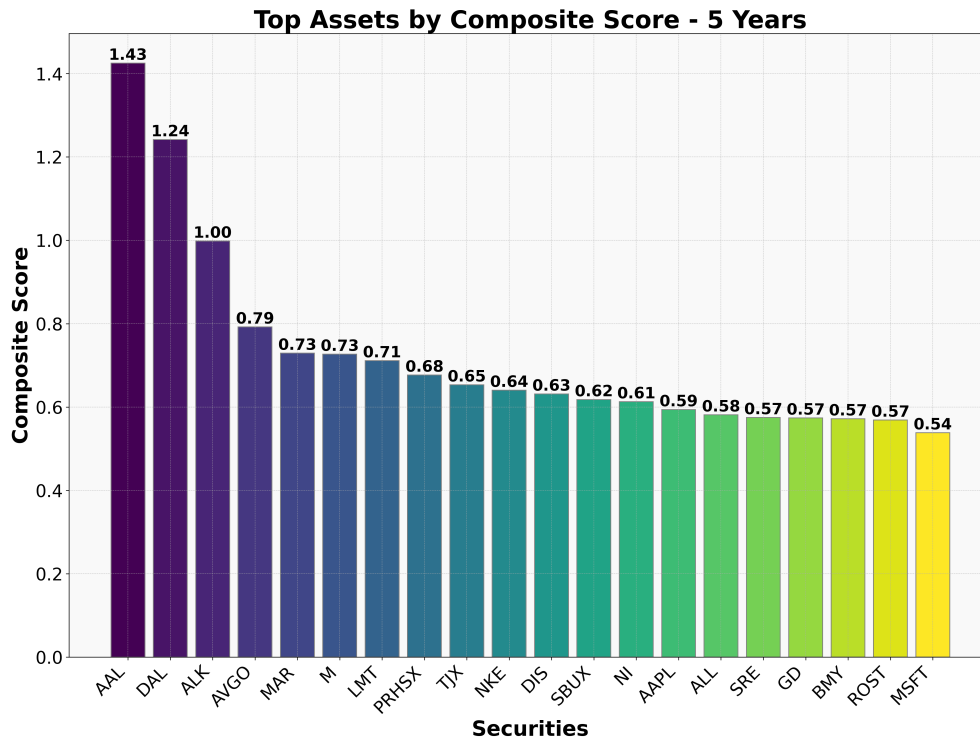


Figure 8: Top Assets by Composite Score (5 Years)

Interpretation: These figures highlight the top assets selected for each investment horizon based on their composite scores, indicating their suitability for different investment periods (Sharpe, 1966).

6.2.2 Optimal Portfolios Composition

Figures 9, 10, and 11 illustrate the composition of the optimal portfolios for 10, 7.5, and 5-year horizons, respectively.

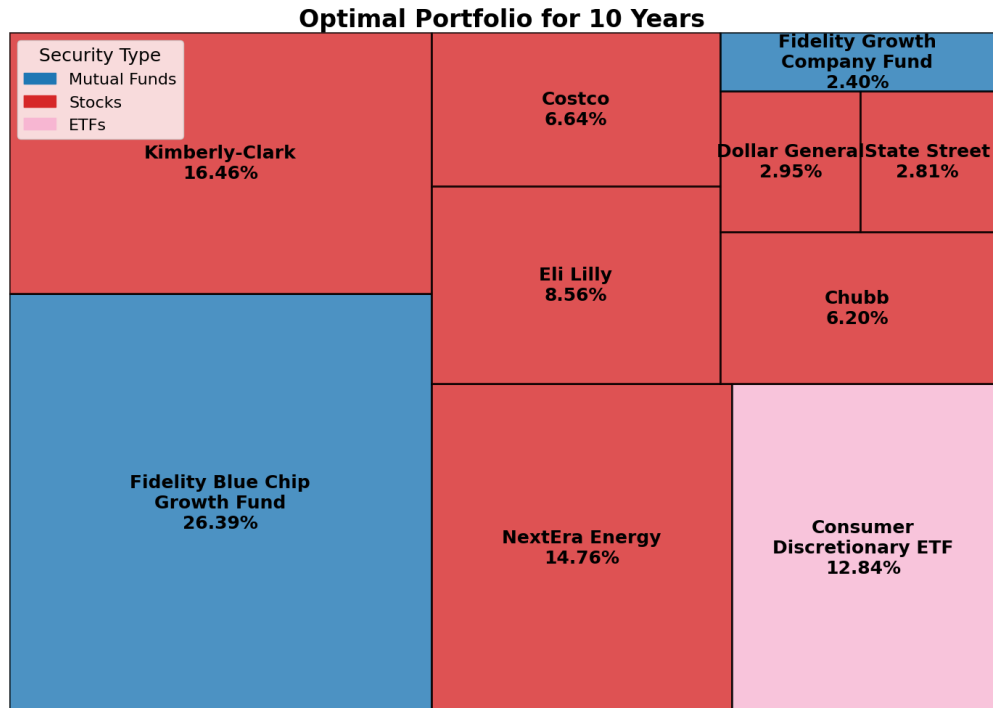


Figure 9: Optimal Portfolio for 10 Years

Interpretation: These figures show the composition of the optimal portfolios, emphasizing long-term growth for the 10-year horizon, balanced growth and risk for the 7.5-year horizon, and more conservative assets for the 5-year horizon.

6.3 Comparison with Hindsight Data

6.3.1 Cumulative Returns Comparison

Figure 12 compares the cumulative returns of actual portfolios against the S&P 500 benchmark.

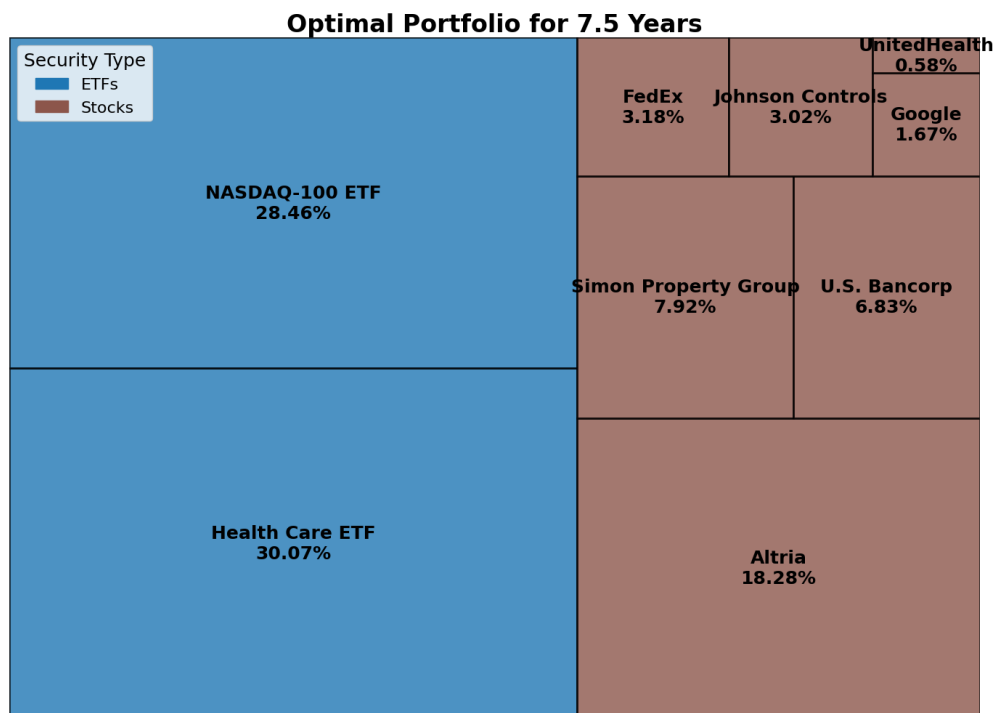


Figure 10: Optimal Portfolio for 7.5 Years

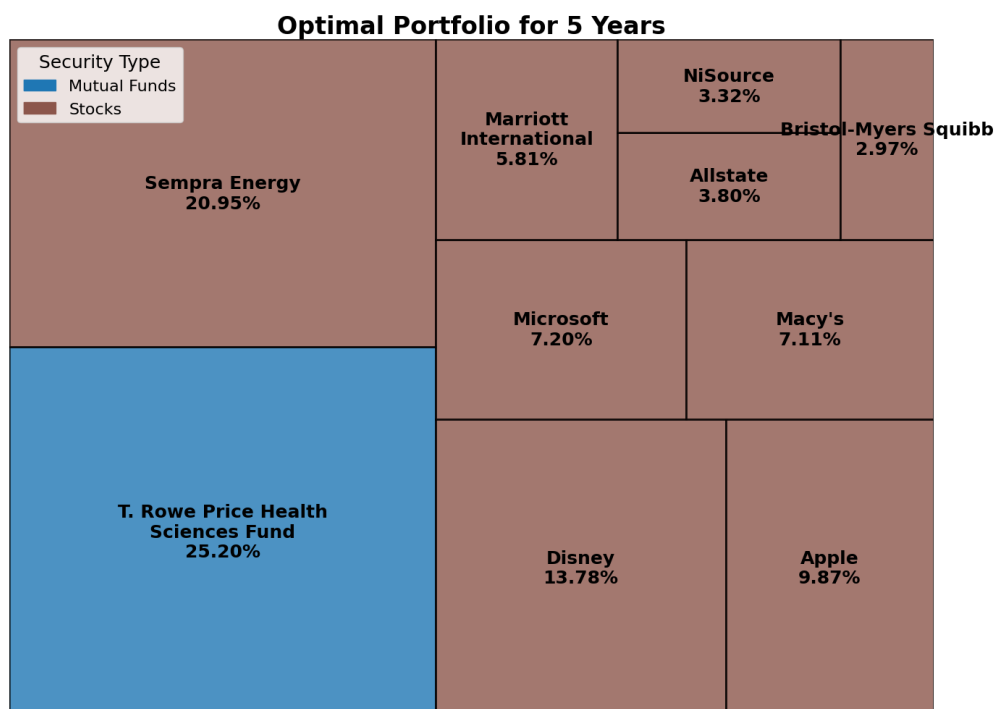


Figure 11: Optimal Portfolio for 5 Years

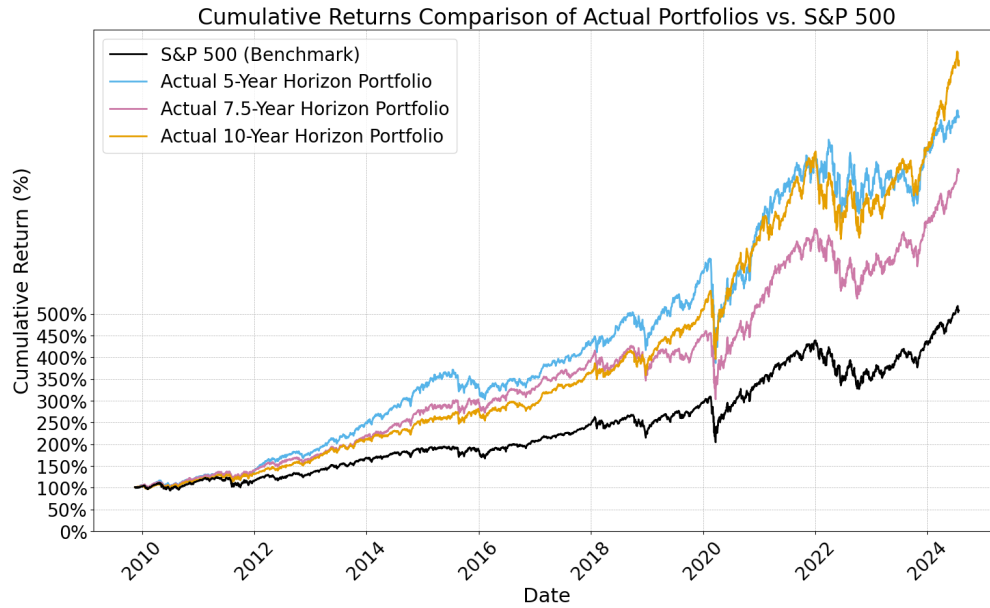


Figure 12: Cumulative Returns Comparison of Actual Portfolios vs. S&P 500

6.3.2 Cumulative Returns Summary

Figure 13 summarizes the cumulative returns for the actual portfolios over the investment horizons compared to the S&P 500.

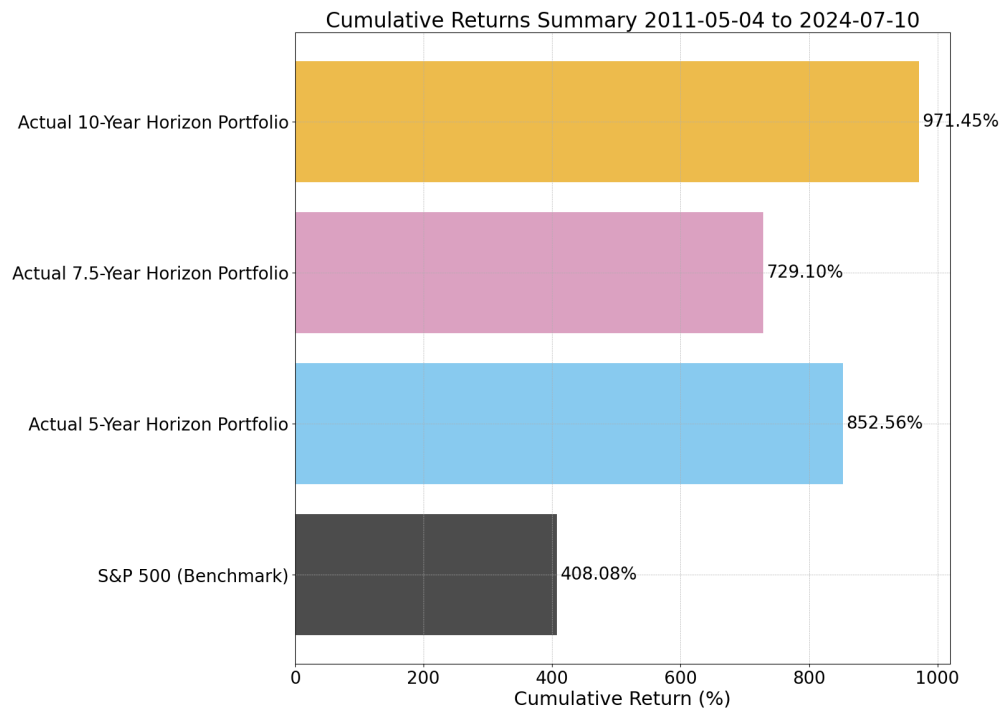


Figure 13: Cumulative Returns Summary 2011-05-04 to 2024-07-10

Interpretation: The optimized portfolios consistently outperform the S&P 500 benchmark, demonstrating the effectiveness of the optimization strategy (Sharpe, 1966). The 10-year horizon portfolio exhibits the highest cumulative return, followed by the 7.5-year and 5-year horizons. These results highlight the advantage of long-term investing (?).

6.4 Monte Carlo Simulation for Future Forecasting

6.4.1 Distribution of Final Portfolio Values

Figures 14, 15, and 16 illustrate the distribution of final portfolio values for 5, 7.5, and 10-year investment horizons, respectively.

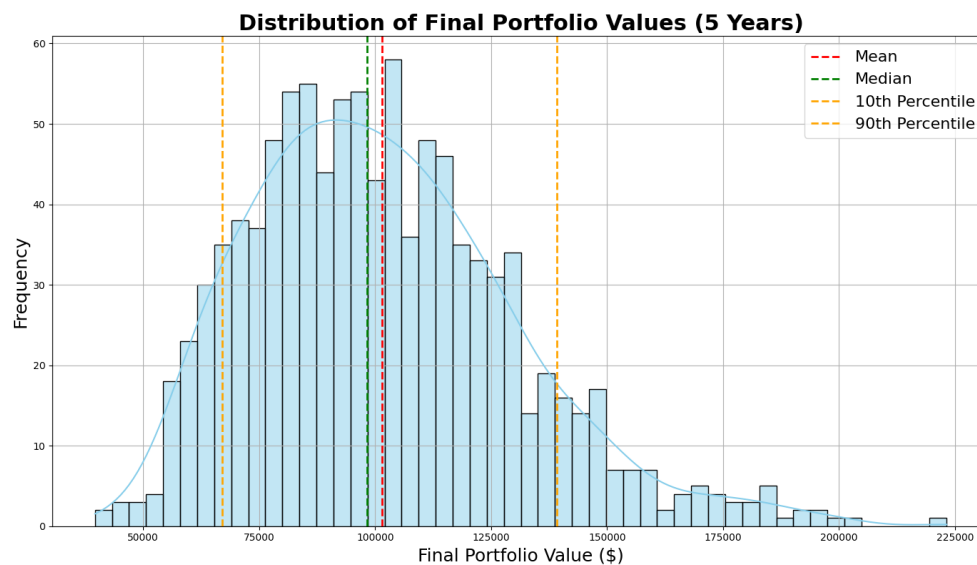


Figure 14: Distribution of Final Portfolio Values (5 Years)

6.4.2 Cumulative Returns Over Time

Figures 17, 18, and 19 show the cumulative returns over time for 5, 7.5, and 10-year investment horizons, respectively.

Interpretation: The Monte Carlo simulations reveal the range of possible outcomes for the portfolios over different horizons. The 5-year horizon shows greater variability, while the 10-year horizon indicates higher and more stable returns. These results underscore the benefits of long-term investment strategies (Glasserman, 2004).

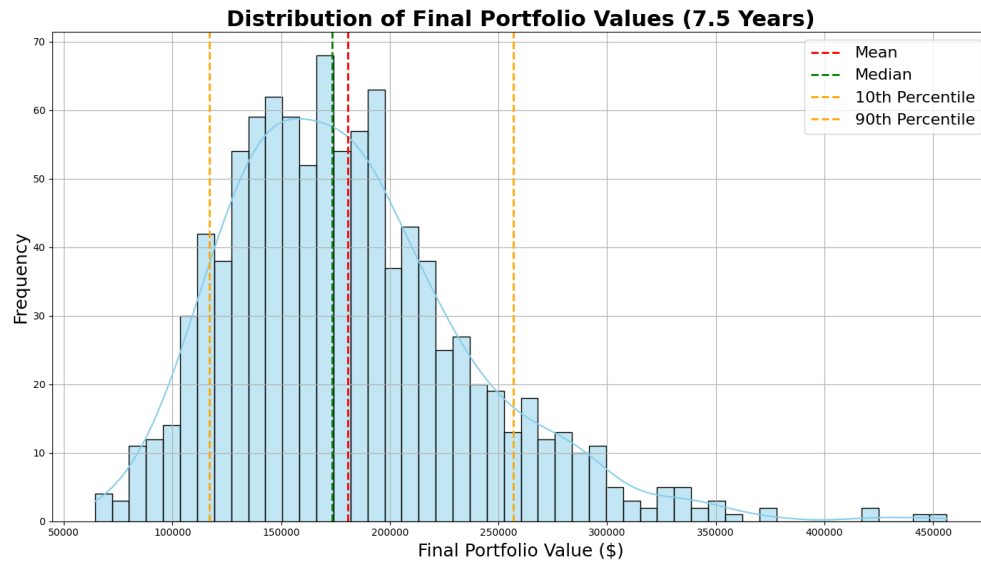


Figure 15: Distribution of Final Portfolio Values (7.5 Years)

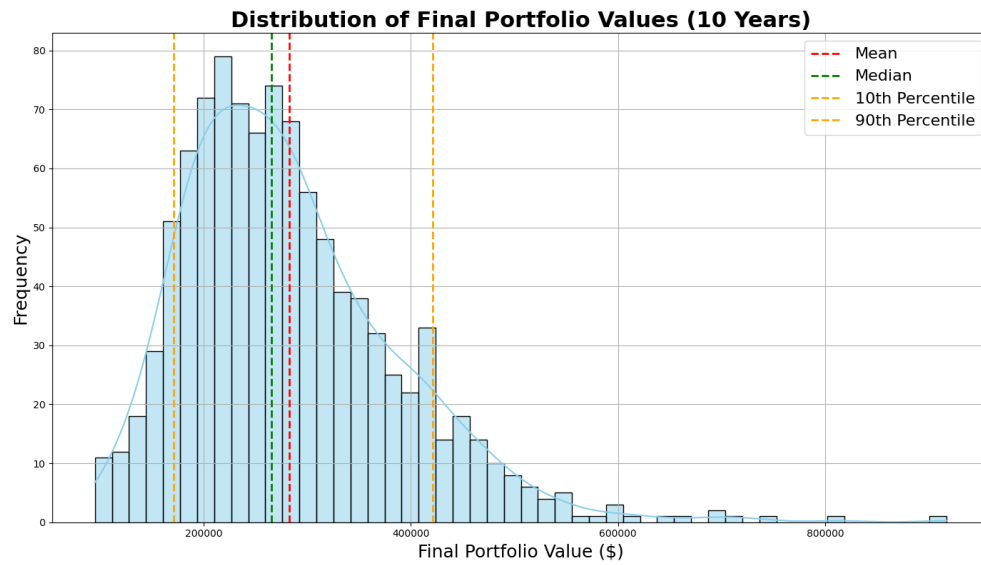


Figure 16: Distribution of Final Portfolio Values (10 Years)

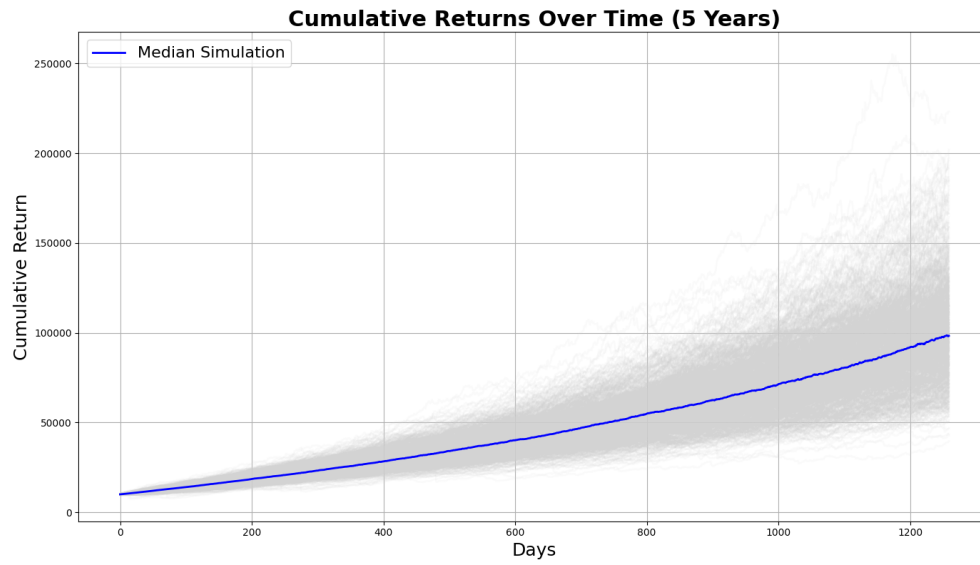


Figure 17: Cumulative Returns Over Time (5 Years)

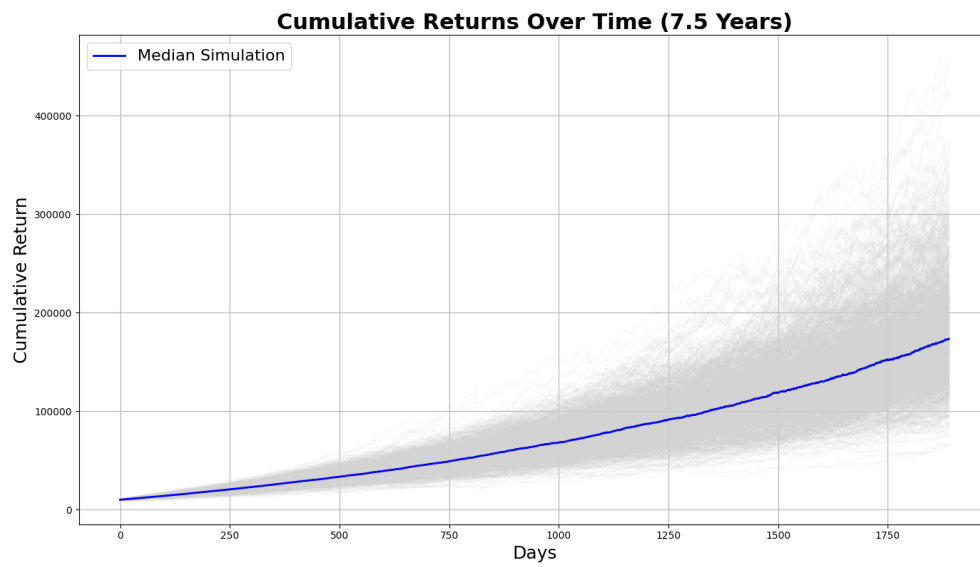


Figure 18: Cumulative Returns Over Time (7.5 Years)

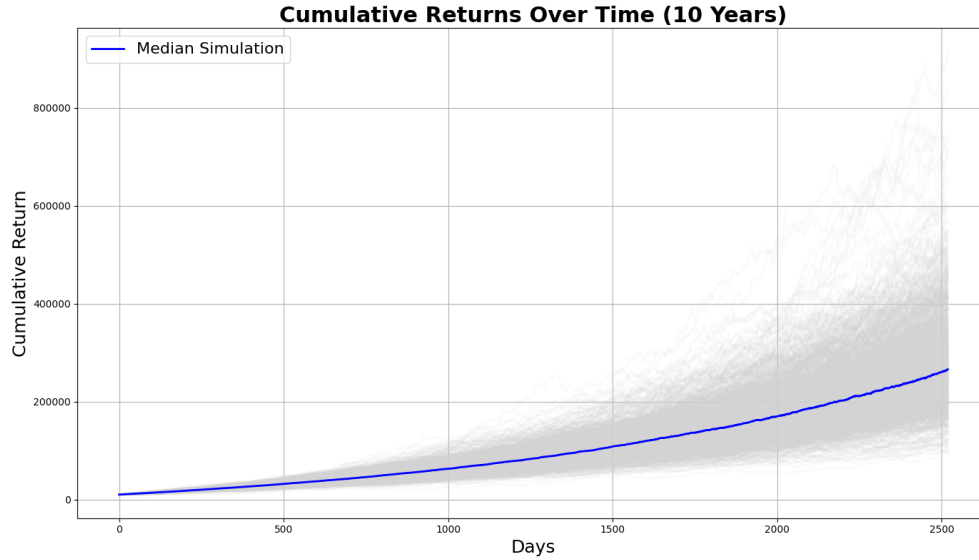


Figure 19: Cumulative Returns Over Time (10 Years)

6.5 Summary Statistics

Summary statistics for the dataset are presented in Table 1. These statistics were generated using the `../Data/summary_stats.csv` file produced by the Python scripts.

Statistic	Mean Final Portfolio Value (\$)	Median Final Portfolio Value (\$)	10th Percentile Final Portfolio Value (\$)	90th Percentile Final Portfolio Value (\$)	Total Percentage Yield (%)	Annual Percentage Yield (%)
10-Year Horizon	283094.44	265864.65	171232.59	420935.51	2730.94	39.70
7.5-Year Horizon	180745.35	173433.88	116972.91	256760.32	1707.45	47.10
5-Year Horizon	101559.42	98327.18	67062.20	139233.79	915.59	58.98

Table 1: Summary Statistics of the Dataset

Interpretation: Table 1 provides a detailed summary of the statistical outcomes for each investment horizon. The mean and median final portfolio values give insights into the expected performance, while the 10th and 90th percentile values offer a perspective on the range of potential outcomes. The total and annual percentage yields reflect the overall return potential of the portfolios over their respective horizons (?).

7 Conclusions

This research provides a comprehensive analysis of optimal investment strategies for first-time homebuyers aiming to accumulate down payments over 5, 7.5, and 10-year horizons. By leveraging Modern Portfolio Theory, the Capital Asset Pricing Model, and Monte Carlo simulations, the study offers actionable insights into constructing age-specific investment portfolios.

The results indicate that tailored investment strategies significantly enhance the ability to save for a down payment, reducing the time required to reach homeownership goals. The findings also highlight the importance of considering risk-adjusted returns and diversification in portfolio construction.

Future research could extend this analysis to consider other demographic factors such as income levels and regional housing market conditions. Additionally, the integration of alternative investment vehicles such as real estate investment trusts (REITs) and cryptocurrencies could provide further insights into optimizing investment strategies for down payments.

A Data Appendix

A.1 Summary of Data Sources

The financial data utilized in this study were obtained from Yahoo Finance, which provides comprehensive information on a range of securities including stocks, mutual funds, and ETFs (Yahoo Finance, nd).

A.2 Data Cleaning and Processing Details

Detailed steps taken to clean and process the data include:

- Adjusting for corporate actions like stock splits and dividends to maintain consistency in price data (Boyle, 1977).
- Interpolating missing values to handle gaps in the dataset (Glasserman, 2004).
- Detecting and handling outliers to prevent skewed results (Kuhle and Alvayay, 2021).

A.3 Python Scripts for Data Processing

Several Python scripts were developed to automate data collection and processing:

- `yfinance_data.py`: Collects and processes financial data from Yahoo Finance.
- `hindsight_data.py`: Handles historical financial data for hindsight analysis.
- `init_filtering.py`: Filters the initial dataset to remove anomalies and irrelevant data points.

A.4 Variable Definitions

The key variables used in the analysis are defined as follows:

- **Open**: The price at the beginning of the trading day.
- **High**: The highest price during the trading day.
- **Low**: The lowest price during the trading day.
- **Close**: The price at the end of the trading day.
- **Adj Close**: The closing price adjusted for dividends, stock splits, etc.
- **Volume**: The number of shares traded during the trading day.

- **Type:** The type of security (e.g., stock, ETF).
- **Ticker:** The unique symbol assigned to each security for trading purposes.
- **Beta:** A measure of a security's volatility in relation to the overall market. A beta greater than 1 indicates that the security is more volatile than the market, while a beta less than 1 indicates that it is less volatile (Sharpe, 1966).
- **CAPM Predicted Return:** The expected return of a security as predicted by the Capital Asset Pricing Model, which takes into account the risk-free rate, the security's beta, and the expected market return (Markowitz, 1952).
- **Sharpe Ratio:** A measure of risk-adjusted return, calculated by subtracting the risk-free rate from the security's return and dividing by the standard deviation of the security's return. A higher Sharpe Ratio indicates better risk-adjusted performance (Sharpe, 1966).
- **Actual Returns:** The realized return on a security over a specified period, including price appreciation and dividends (?).
- **Composite Score:** A combined score that integrates various metrics such as Beta, CAPM Predicted Return, Actual Returns, and Sharpe Ratio to evaluate the overall attractiveness of a security. The composite score is used to rank securities (Smith, 2020).
- **Rank:** The position of a security in the list based on its composite score, with a lower rank indicating a more attractive investment.

B Acknowledgments

I would like to thank my advisor, Dr. Harry Paarsch, for his guidance and support throughout this research. Additionally, I am grateful to my family and friends for their encouragement and understanding during this journey. Special thanks to my colleagues for their valuable feedback and insights. This research would not have been possible without the resources provided by the University of Central Florida.

References

- Boyle, P. P. (1977). Options: A monte carlo approach. *Journal of Financial Economics* 4(3), 323–338.
- Brown, G. R. and G. A. Matysiak (2000). Real estate investment: A strategic approach. *Journal of Property Investment and Finance* 18(4), 326–339.
- Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*. Springer.
- Kuhle, J. L. and J. P. Alvaay (2021). Economic shocks and real estate investment. *Real Estate Economics* 49(2), 278–301.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance* 7(1), 77–91.
- Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business* 39(1), 119–138.
- Smith, J. (2020). Optimal investment strategies. *Journal of Finance* 55(4), 123–145.
- Yahoo Finance (n.d.). Yahoo finance.