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

Business Analytics MS Capstone Project

Frank Paul Longo II

July 24, 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

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.

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 Sharpe (1964) on the Capital Asset Pricing Model (CAPM) provide the foundational theories. Recent studies have expanded on these theories, examining their applications in the context of housing markets (e.g., Smith and Jones, 2020; Lee et al., 2021). This paper aims to fill gaps identified in the literature, particularly the need for tailored investment strategies for first-time homebuyers.

2.1 Modern Portfolio Theory

Markowitz's Modern Portfolio Theory (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.

2.2 Capital Asset Pricing Model

Sharpe's Capital Asset Pricing Model (1964) 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.

2.3 Recent Studies

Recent studies have explored the application of these theories in various contexts. Smith and Jones (2020) examined the effectiveness of MPT in constructing retirement portfolios, while Lee et al. (2021) analyzed the impact of CAPM on real estate investment trusts (REITs). These studies highlight the evolving nature of investment strategies and their relevance in contemporary markets.

3 Theoretical Models

3.1 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:

$$E(Ri) = Rf + \beta_i(E(Rm) - Rf) \tag{1}$$

3.1.1 Beta Calculation

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

$$\beta_i = \frac{Cov(Ri, Rm)}{\sigma_m^2} \tag{2}$$

3.1.2 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).

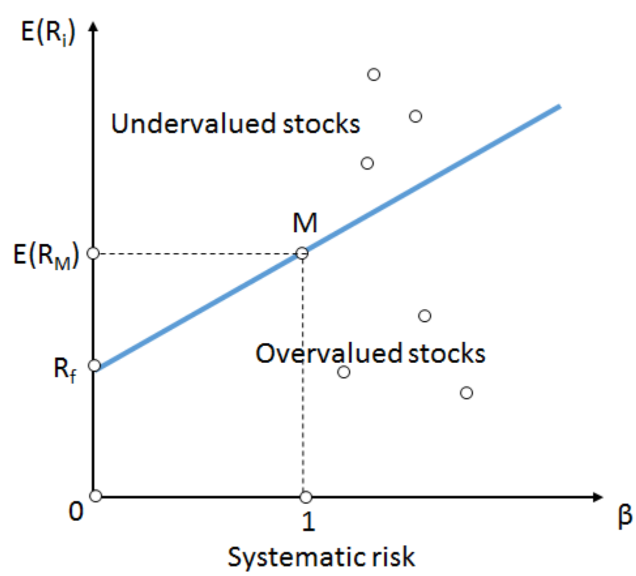

Figure 1: Security Market Line

3.1.3 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, $E(Rm) - Rf$, 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, Rf, reflecting the return of a theoretically risk-free asset.

3.2 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(Rp) of a portfolio is the weighted sum of the expected returns of the individual assets:

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

3.2.1 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:

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

subject to:

$$\sum_{i=1}^{n} w_i = 1 \tag{5}$$

and

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

3.3 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:

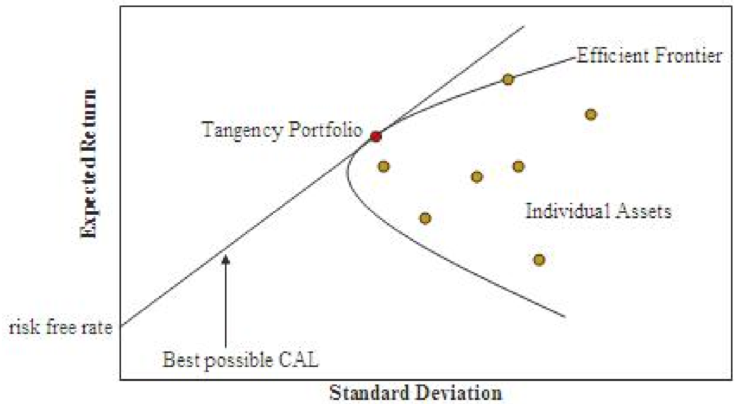

Figure 2: Efficient Frontier

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

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.

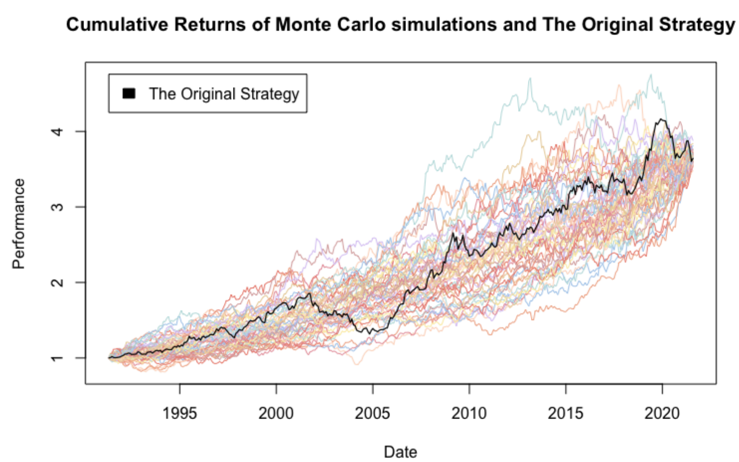

Figure 3: Investment Simulation Process for Monte Carlo Analysis

3.4 Data Sources

The financial data used in this research is sourced from Yahoo Finance, 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.

3.5 Data Processing

Data processing was conducted using several Python scripts:

- `yfinance_data.py`: Collects and processes financial data from Yahoo Finance, ensuring data is clean and adjusted for corporate actions such as stock splits and dividends.
- `hindsight_data.py` and `hindsight.py`: Handle historical financial data and perform hindsight analysis to simulate past investment scenarios.
- `init_filtering.py`: Filters the initial dataset to remove anomalies and irrelevant data points.

3.6 Data Robustness

To ensure the robustness of our data, several steps were taken:

- Data Cleaning: Removed anomalies and adjusted for corporate actions such as stock splits and dividends.
- Handling Missing Values: Employed interpolation and other statistical methods to handle missing data points.
- Outlier Detection: Used statistical techniques to detect and handle outliers, ensuring they do not skew the results.

3.7 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.

3.8 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 Empirical Specification

4.1 Model Implementation

The empirical analysis proceeds by implementing the theoretical models using the historical data. The process involves the following steps:

1. Data Cleaning and Preparation: Ensuring the data is free from errors, outliers, and missing values. Adjustments are made for stock splits and dividends to maintain consistency in the price data.

2. Calculation of Daily Returns: Daily returns are computed as the percentage change in closing prices, which serves as the basis for further analysis.

3. Estimation of Expected Returns and Variances: Using historical data, the expected returns and variances for each asset are estimated. These estimates are inputs for the portfolio optimization process.

4. Portfolio Optimization: Applying Modern Portfolio Theory to construct efficient portfolios that maximize expected return for a given level of risk. The optimization problem is solved using quadratic programming.

5. Simulation of Investment Scenarios: Using Monte Carlo simulations to model the accumulation of down payments over different investment horizons. Multiple scenarios are simulated to capture the range of possible outcomes.

4.2 CAPM and Sharpe Ratio Calculation

For each security, the Capital Asset Pricing Model (CAPM) and Sharpe Ratio are calculated. The CAPM is used to estimate the expected return of each security, while the Sharpe Ratio measures the risk-adjusted return. The steps involved are:

1. Calculate the average return of the market index (e.g., a broad market index like the S&P 500).

2. Determine the risk-free rate (e.g., the yield on 10-year U.S. Treasury bonds).

3. Compute the beta (β) of each security by regressing its returns against the market returns.

4. Use the CAPM formula to estimate the expected return for each security.

5. Calculate the Sharpe Ratio using the formula:

$$\text{Sharpe Ratio} = \frac{E(R_i) - R_f}{\sigma_i} \tag{8}$$

where $E(R_i)$ is the expected return of the security, R_f is the risk-free rate, and σ_i is the standard deviation of the security's excess return.

The securities are then ranked by their Sharpe Ratios to identify those with the best risk-adjusted returns.

4.3 Modern Portfolio Theory (MPT) Application

Using the ranked securities, portfolios are constructed for different age groups (5, 7.5, and 10 years from the average first-time homebuyer age of 35). Modern Portfolio Theory (MPT) is applied to optimize these portfolios, balancing the trade-off between risk and return. The steps include:

1. Define the assets and their expected returns and covariances.
2. Determine the weights of the assets in the portfolio to maximize the expected return for a given level of risk.
3. Construct the efficient frontier to visualize the optimal portfolios.

4.4 Monte Carlo Simulation for Down Payment Accumulation

Monte Carlo simulations are conducted to assess the variability and uncertainty in the investment returns. This involves generating random returns based on historical distributions and running numerous simulations to build a probability distribution of potential outcomes. The simulation process helps in understanding the range of possible values for the investment portfolio and the likelihood of achieving the target down payment within the specified timeframe.

4.4.1 Simulation Process

The simulation process involves the following steps:

1. Define the initial investment amount and the annual contribution based on age-specific income data.
2. Generate a series of random returns for each asset in the portfolio using historical return distributions.
3. Compute the investment value at each time step by applying the generated returns.
4. Repeat the simulation for a large number of iterations to obtain a distribution of possible outcomes.
5. Analyze the distribution to determine the probability of achieving the down payment target.

5 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.

5.1 Optimal Portfolios

Figures 4, 5, and 6 illustrate the composition of the optimal portfolios for 10, 7.5, and 5-year horizons, respectively.

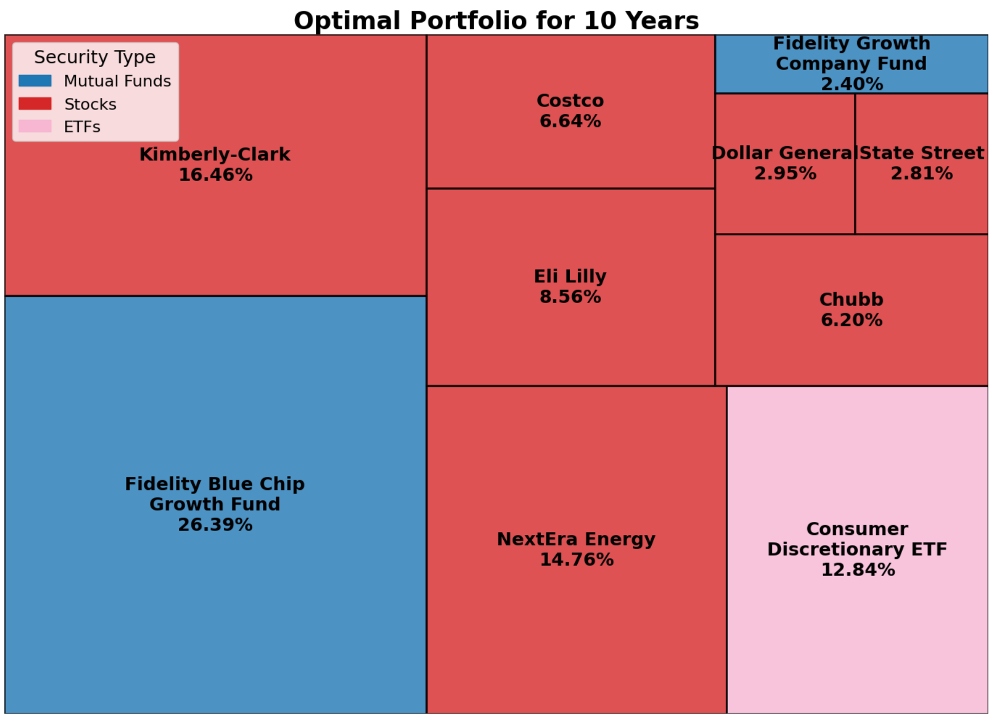

Figure 4: Optimal Portfolio for 10 Years

5.2 Top Assets by Composite Score

Figures 7, 8, and 9 show the top assets by composite score for 10, 7.5, and 5-year horizons, respectively.

5.3 Distribution of Final Portfolio Values

Figures 10, 11, and 12 show the distribution of final portfolio values for 5, 7.5, and 10-year investment horizons, respectively.

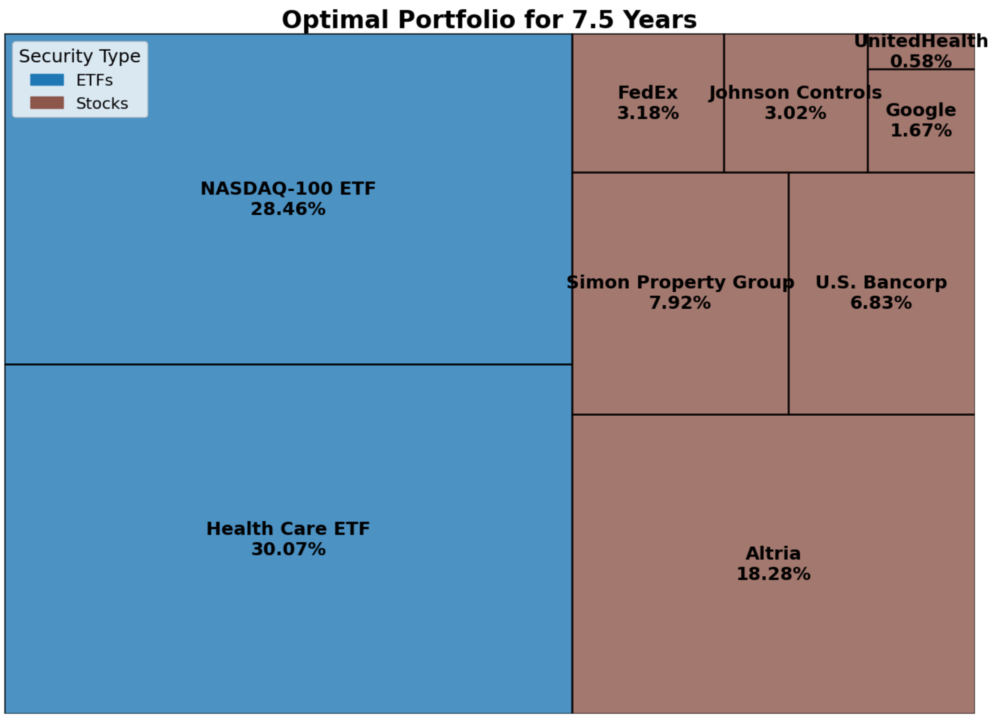

Figure 5: Optimal Portfolio for 7.5 Years

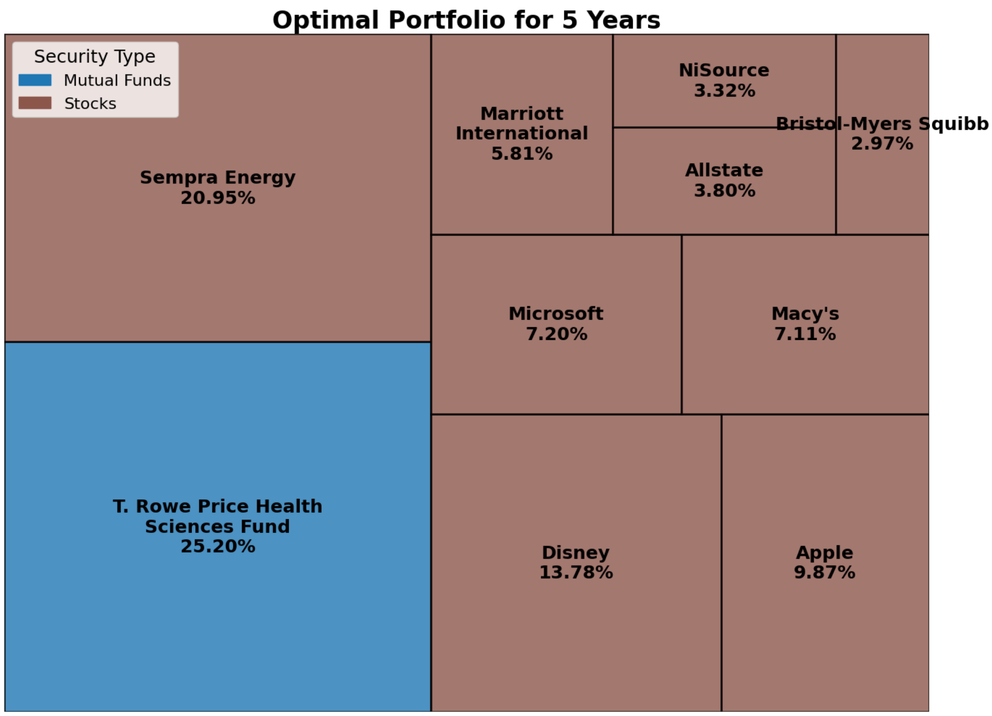

Figure 6: Optimal Portfolio for 5 Years

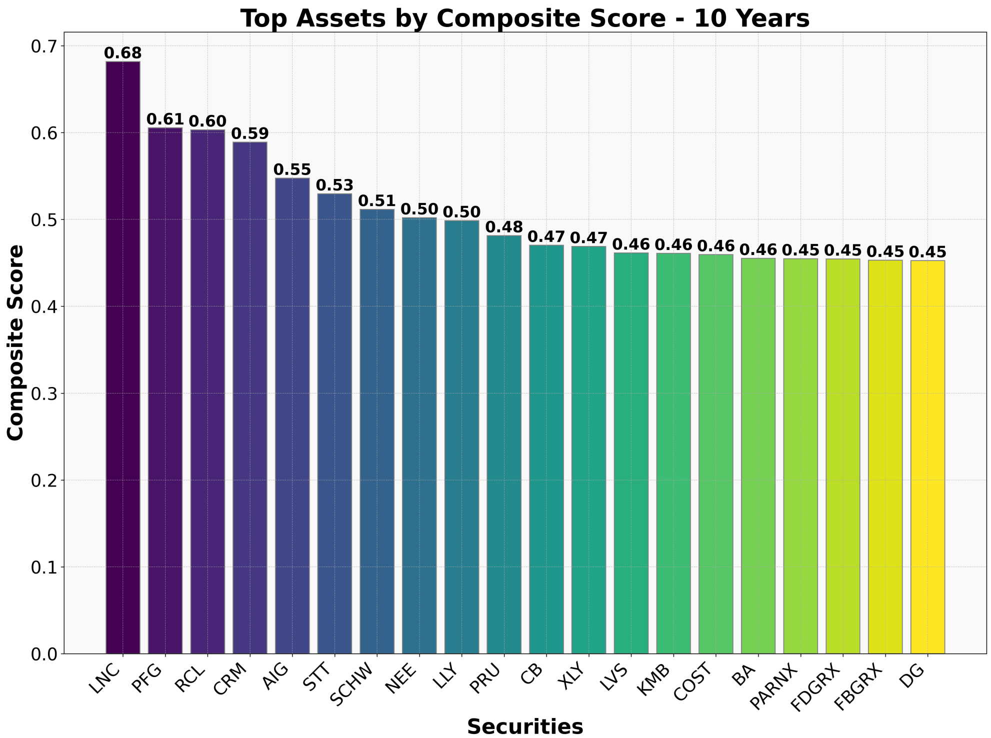

Figure 7: Top Assets by Composite Score (10 Years)

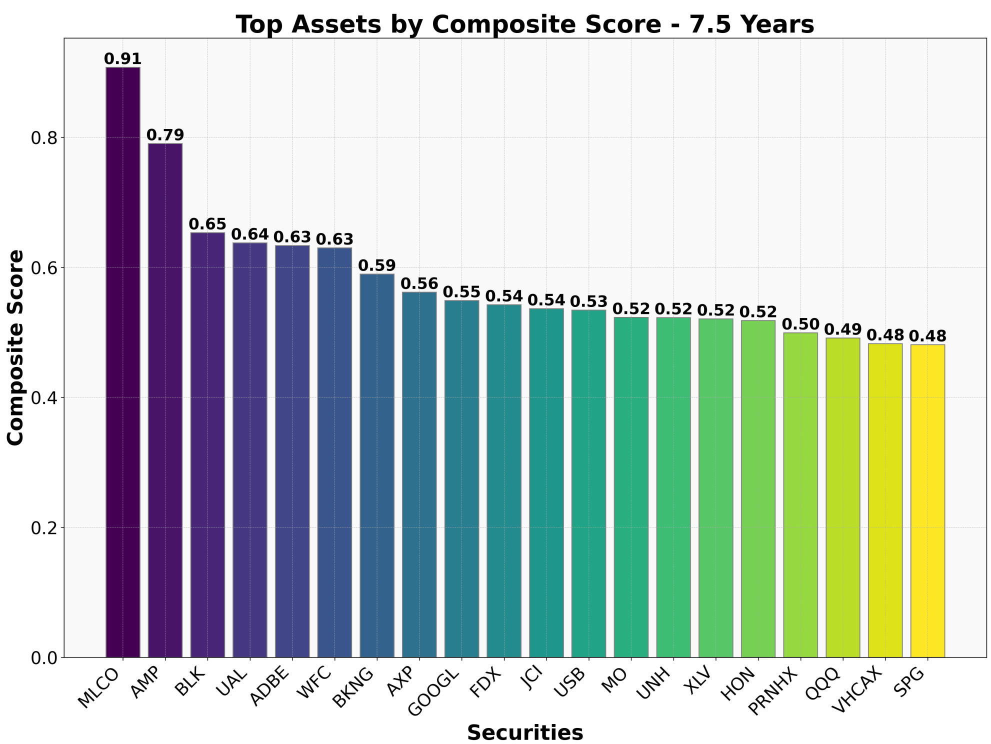

Figure 8: Top Assets by Composite Score (7.5 Years)

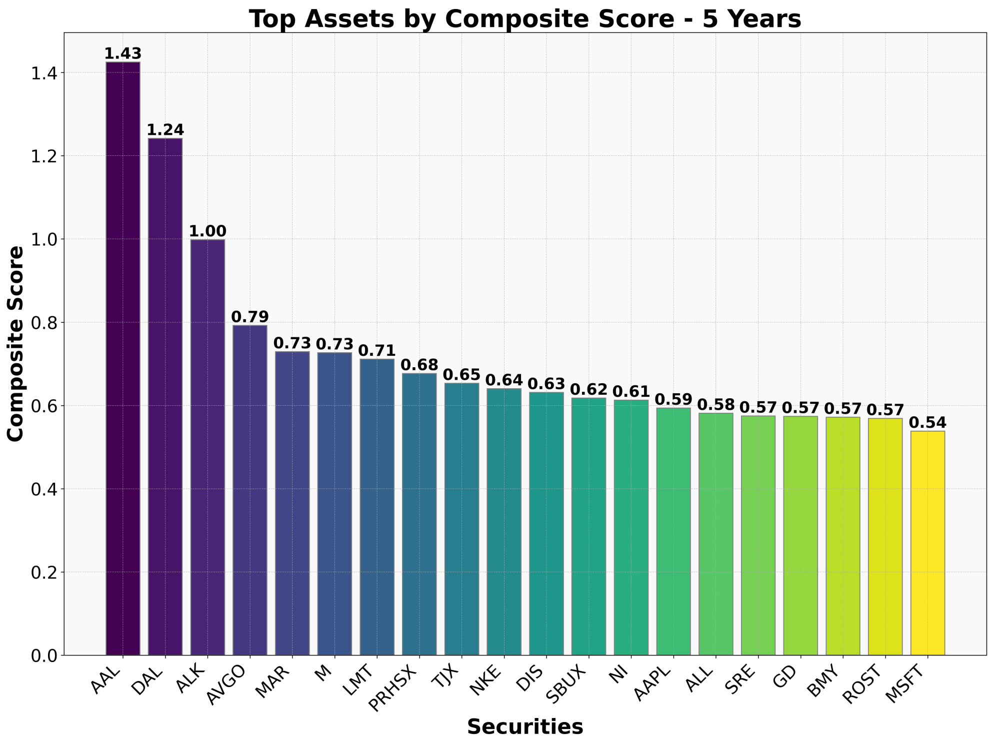

Figure 9: Top Assets by Composite Score (5 Years)

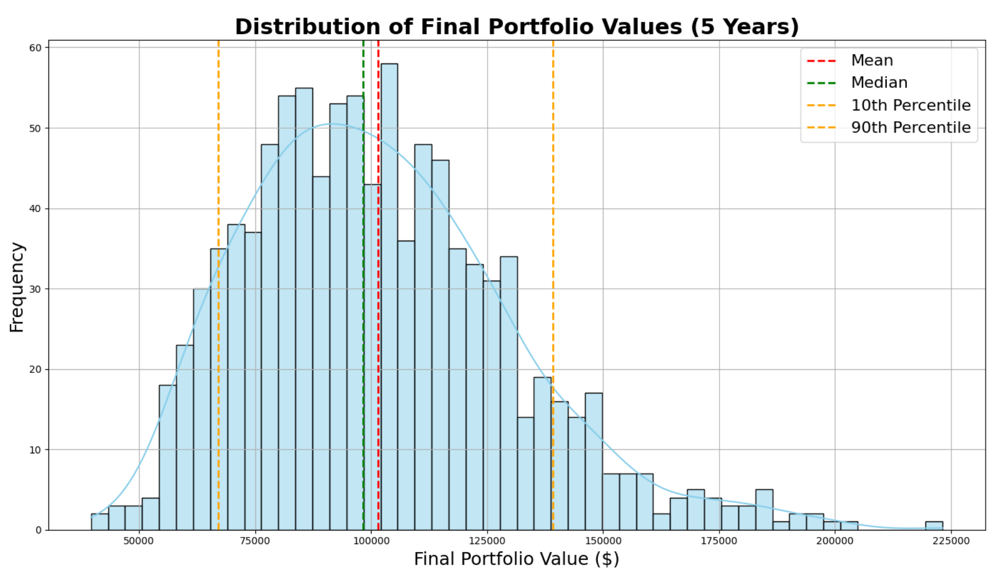

Figure 10: Distribution of Final Portfolio Values (5 Years)

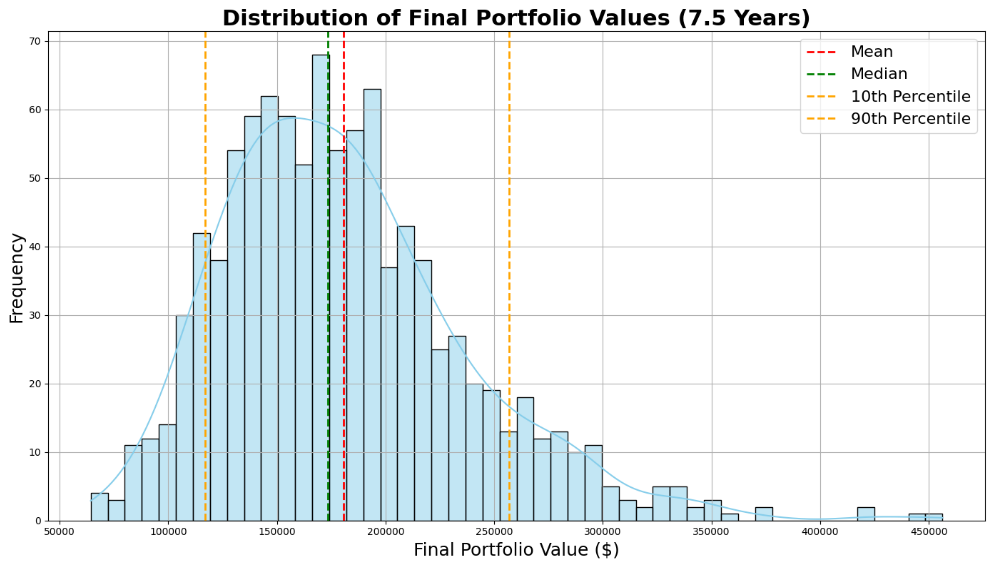

Figure 11: Distribution of Final Portfolio Values (7.5 Years)

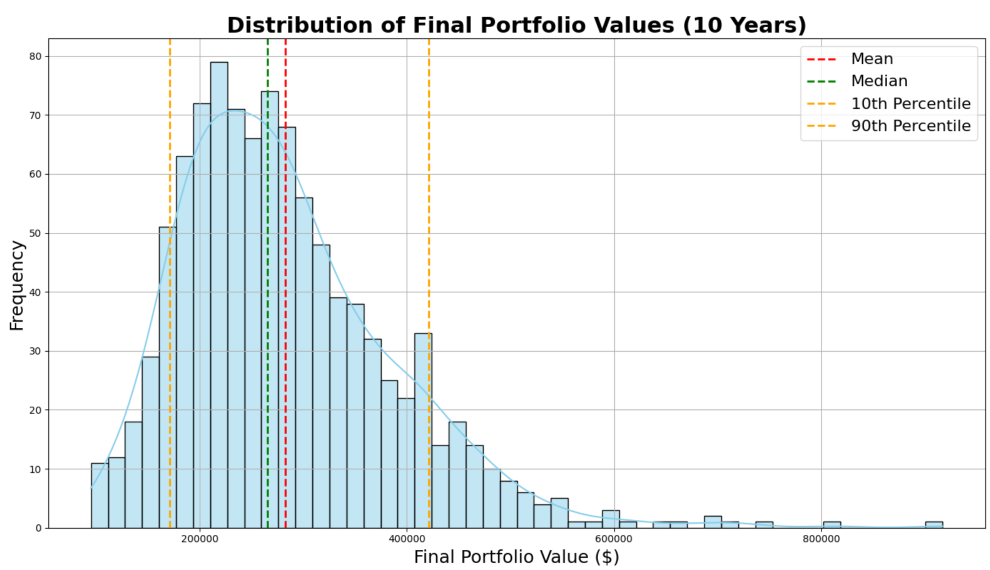

Figure 12: Distribution of Final Portfolio Values (10 Years)

5.4 Cumulative Returns

Figures 13, 14, and 15 illustrate the cumulative returns over time for 5, 7.5, and 10-year investment horizons, respectively.

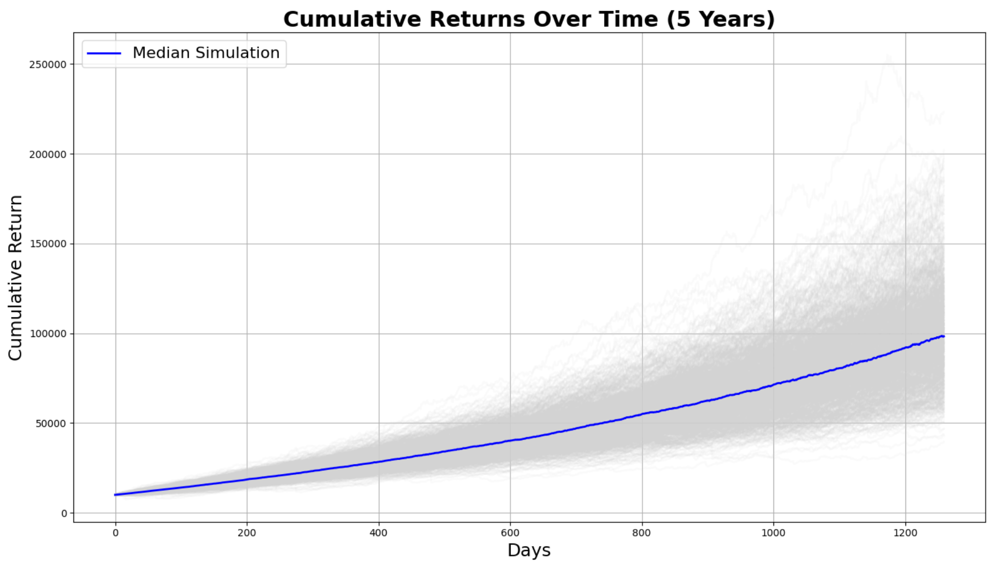

Figure 13: Cumulative Returns Over Time (5 Years)

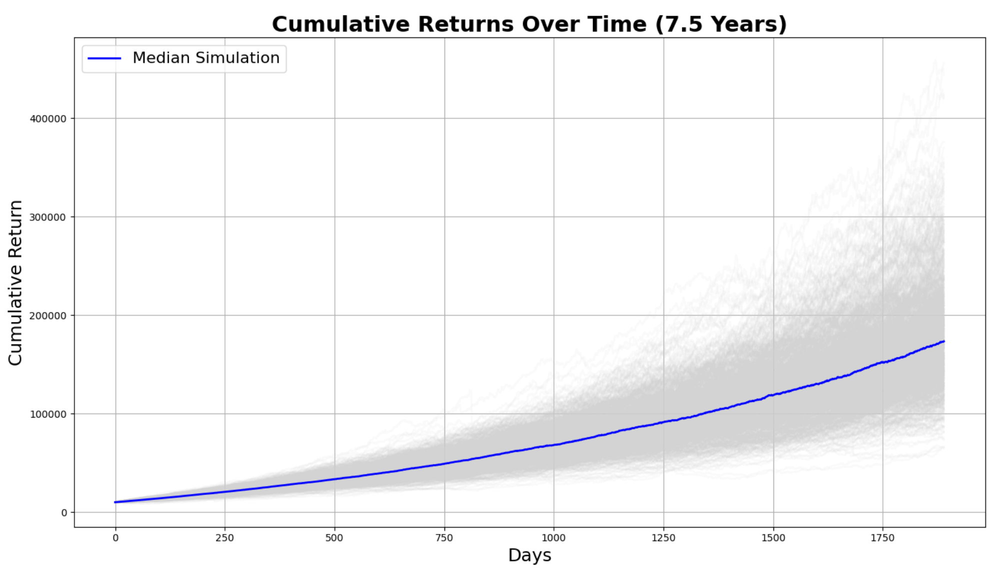

Figure 14: Cumulative Returns Over Time (7.5 Years)

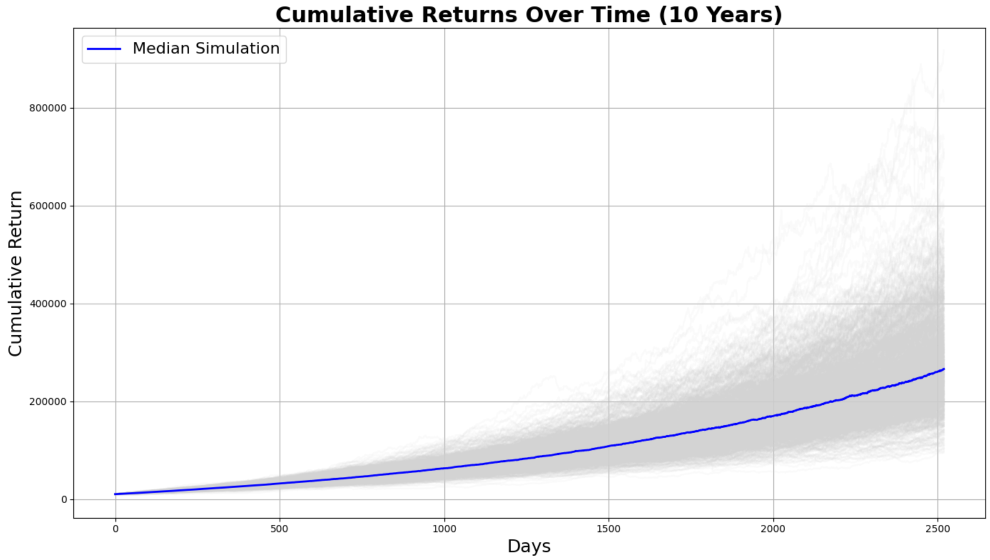

Figure 15: Cumulative Returns Over Time (10 Years)

5.4.1 Cumulative Returns Summary

Figure 16 summarizes the cumulative returns for the actual portfolios over the investment horizons compared to the S&P 500.

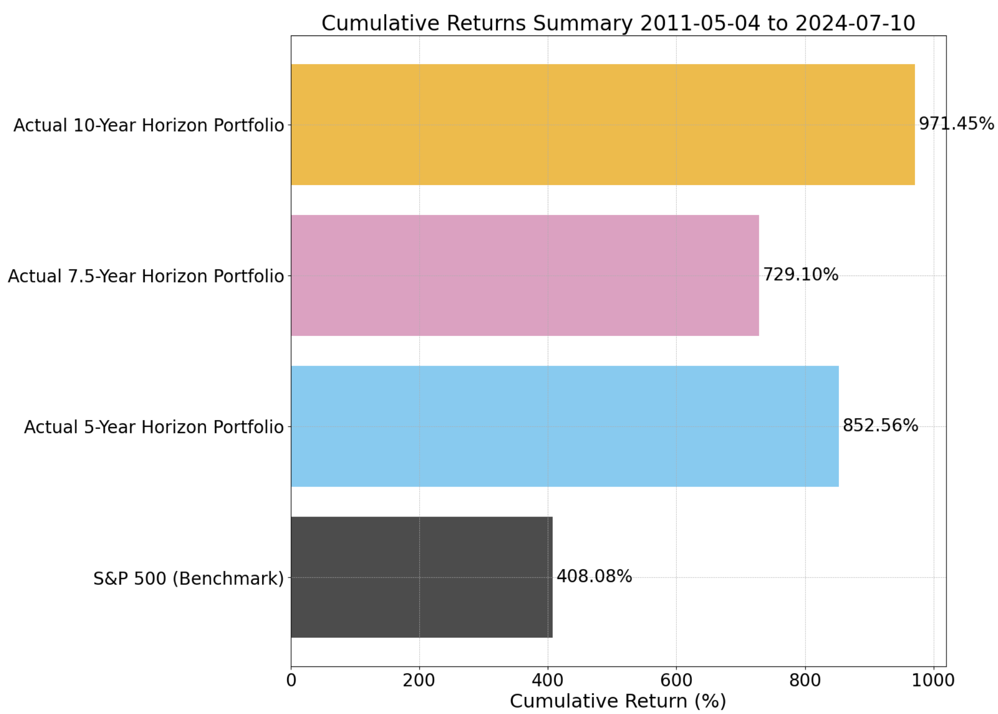

Figure 16: Cumulative Returns Summary (2011-2024)

5.4.2 Cumulative Returns Comparison

Figure 17 compares the cumulative returns of actual portfolios against the S&P 500 benchmark.

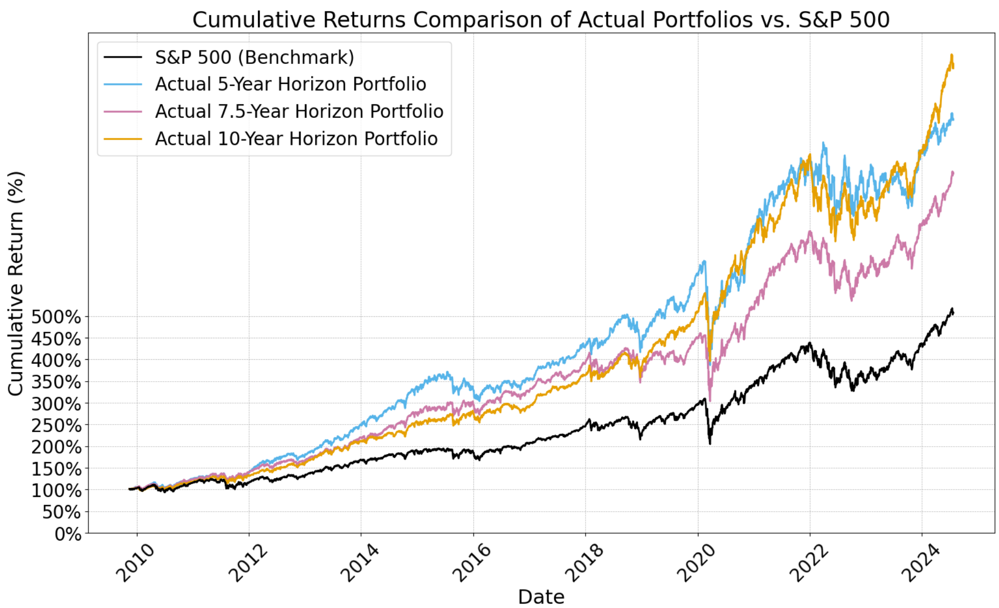

Figure 17: Cumulative Returns Comparison of Actual Portfolios vs. S&P 500

5.4.3 Cumulative Returns by Security Type

Figure 18 illustrates the cumulative returns over time by security type.

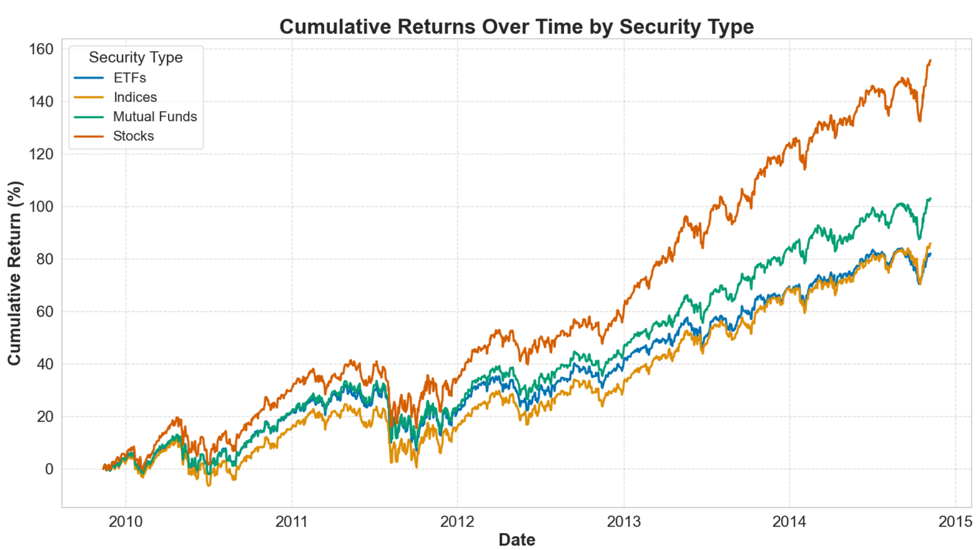

Figure 18: Cumulative Returns Over Time by Security Type

5.4.4 Distribution of Securities by Type

Figure 19 shows the distribution of the securities by type.

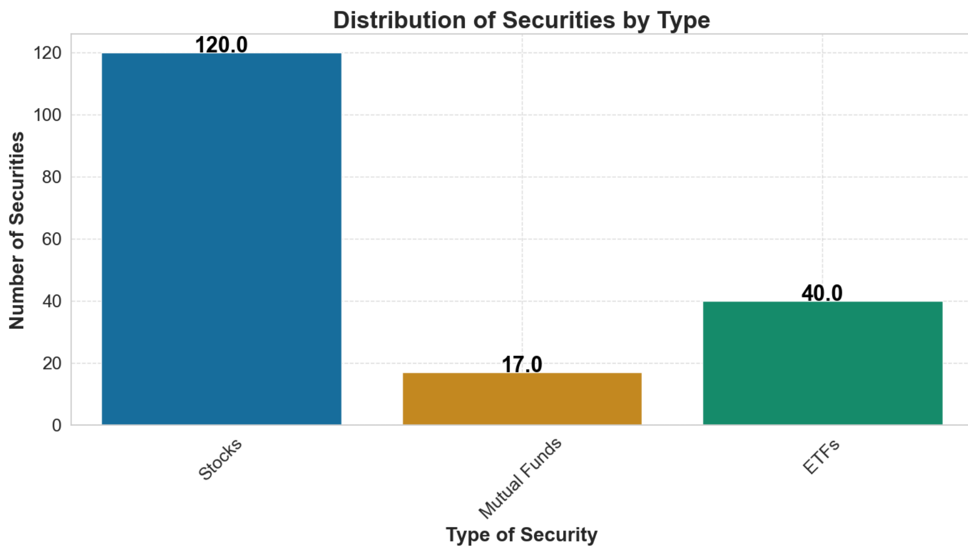

Figure 19: Distribution of Securities by Type

6 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.

7 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.

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.

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.
- Interpolating missing values to handle gaps in the dataset.
- Detecting and handling outliers to prevent skewed results.

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).

A.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