Stock Market Forecasting and Alignment

PIC1

Author: Pedro Maltez

Supervisors: Miguel Couceiro
Vasco Guerra
Ajinkya Kulkarni

Engineering Physics and Technology
Instituto Superior Técnico

2024/2025

Contents

Abstract

This study proposes a data-driven investment framework that dynamically adjusts capital allocation to the SP 500 based on deviations from its long-term exponential growth trend. Unlike traditional "Buy and Hold" strategies that rely on passive compounding, our method systematically identifies periods of market overvaluation and undervaluation, enabling more informed capital deployment. Back tested over a 40-year period (1985–2025), the strategy demonstrates improved terminal wealth compared to passive investing, with Monte Carlo simulations validating its robustness across various market regimes.

1 Introduction

In the last decades, investing in financial markets became increasingly accessible to the general public, leading to a surge in interest around strategies aimed at optimizing long-term returns. Among these, the 'Buy and Hold' [1] strategy has been widely adopted due to its simplicity and historical success. However, this passive approach often overlooks dynamic market conditions and fails to take advantage of opportunities to maximize returns based on the market's current valuation.

The objective of this project was to explore a more dynamic and practical investment strategy, using historical data of the S&P 500 index, particularly focusing on identifying when the market is undervalued or overvalued. By analyzing the historical performance of the index, the project aimed to determine favorable investment periods when the market is undervalued, and avoid overvalued periods. A dynamic allocation of capital will be applied based on these findings, with the goal of improving long-term returns compared to a static 'Buy and Hold' approach.

Additionally, this paper explores the effectiveness of our proposed strategy by simulating the portfolio's performance, using Monte Carlo simulations to assess its potential outcomes and compare the ROI for different scenarios. This approach is designed to help investors by enabling them to capitalize on good investment periods while avoiding unfavorable conditions.

2 Historical Data Analysis

To identify favorable investment periods and avoid overvalued market conditions, a detailed analysis of historical S&P 500 index data was undertaken. This section outlines the data sources, initial assumptions, and methodological framework used in the analysis.

2.1 Source and First Assumptions

The data analysis was carried out using Python within a Jupyter Notebook environment, which provided an interactive platform for processing, visualizing, and modeling financial data. Historical price data for the S&P 500 index was retrieved via the yfinance API, a widely adopted Python library that enables access to Yahoo Finance datasets [2].

We selected a 40-year time frame, with monthly intervals between each data point. This extended period was chosen to reflect the investment horizon of a typical long-term investor, that is, one who allocates funds on a fixed day each month and aims to optimize their contributions over time. Short-term price fluctuations are less relevant in this context, since the focus is placed on identifying favorable periods for allocation within a long-term strategy.

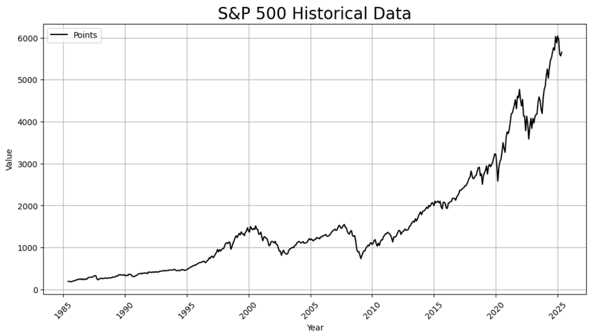

Figure 1: S&P 500 Historical Data

As illustrated in Figure 1, the S&P 500 demonstrates a long-term growth pattern that is approximately exponential, despite intermittent periods of volatility.

2.2 Exponential Trend

To examine the long-term behavior of the S&P 500 index, the historical price data were transformed using a logarithmic scale [3]. This transformation serves to linearize exponential growth patterns, allowing the use of linear regression [4] to estimate the average growth rate over time.

Applying the natural logarithm to the monthly price data and fitting a model of the form

$$\log(P_{\text{S\&P 500}}) = ax + b, \tag{1}$$

The coefficients obtained were approximately 0.0060 and 5.5859, a and b, respectively.

These values were derived through ordinary least squares regression [5], as illustrated in Figure 2:

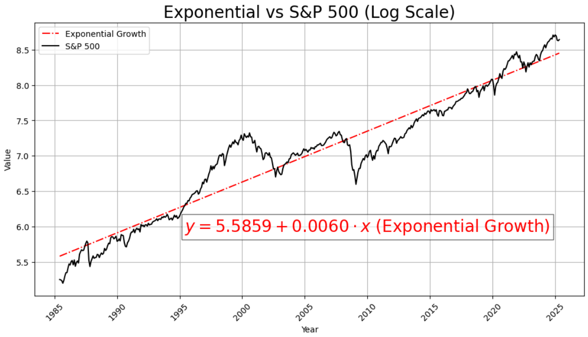

Figure 2: S&P 500 Historical Data in Logarithmic Scale

The original exponential model of the price can be recovered by inverting the logarithmic transformation:

$$P_{\text{S\&P 500}} = e^{ax+b}. \tag{2}$$

The resulting curve represents the average exponential growth trend of the index over the 40-year period and is shown in Figure 3:

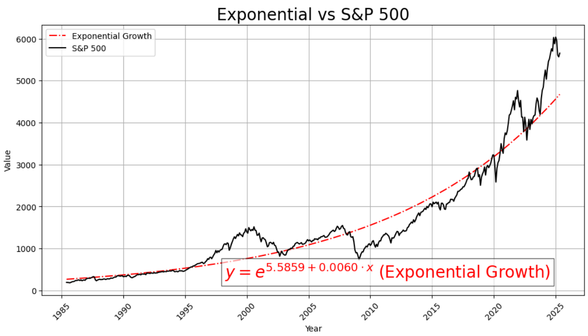

Figure 3: S&P 500 Exponential Growth Approximation

2.3 Compound Annual Growth Rate (CAGR)

To quantify the average annual return over the period, the Compound Annual Growth Rate (CAGR) was calculated using the formula:

$$\text{CAGR} = \left(\frac{P_{\text{final}}}{P_{\text{initial}}}\right)^{\frac{12}{n}} - 1, \tag{3}$$

where P_{final} is the price at the end of the period of n years and P_{initial} is the price at the beginning. Using the average exponential of S&P 500 price data CAGR $= e^{0.0060(12)} - 1 \approx 0.074$, which corresponds to an approximate annual increase of 7.4%.

This confirms that the S&P 500 has experienced an average annual growth of around 7– 8%, validating the use of an exponential model for long-term trend analysis.

2.4 Time to Invest?

With the long-term exponential growth trend of the S&P 500 established, it becomes possible to explore one of the most frequently asked questions in finance: Is now a good time to invest?

A widely accepted strategy, popularized by Warren Buffett, is Buy and Hold, which involves consistently investing a fixed amount at regular intervals, regardless of market conditions. Assuming a historical average monthly return of around 0.6%, this approach would result in a portfolio approximately 585% larger after 40 years, highlighting the power of long-term compounding.

However, this strategy also implies investing both during market highs and lows, exposing the investor to unnecessary risk and missed opportunities. Since the market does not grow uniformly, but fluctuates around the exponential trend, identifying deviations from this baseline may allow for more efficient capital allocation.

This insight motivates the development of a strategy that dynamically adjusts investment amounts based on how overvalued or undervalued the index is relative to its exponential trend.

To capture whether the S&P 500 is overvalued or undervalued relative to its long-term trend, a valuation metric was defined based on the deviation between the actual price and the exponential model:

$$\Delta\Phi(t) = \frac{P_{\text{actual}}(t) - P_{\text{exp}}(t)}{P_{\text{exp}}(t)}, \tag{4}$$

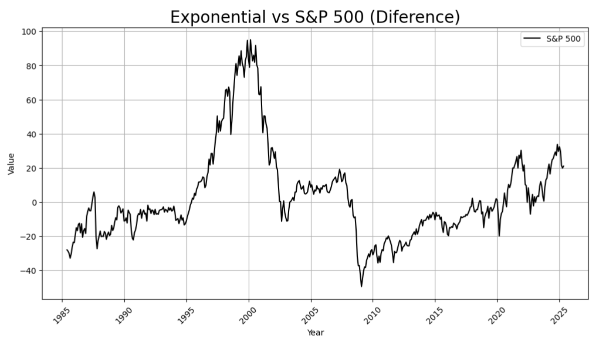

Figure 4: Relative Deviation between Exponential and S&P 500

This expression computes the relative difference between the observed price and the expected price given by the exponential trend at time t. A negative value indicates a potential buying opportunity (undervaluation), while a positive value suggests the market may be overvalued.

By using this valuation metric, it is possible to construct a dynamic investment strategy that increases allocations when the index is undervalued and decreases them when it is overvalued, potentially improving long-term returns and reducing risk compared to a passive Buy and Hold approach.

3 Weighted Investment

This section introduces a dynamic investment strategy that adjusts monthly contributions based on the S&P 500's deviation from its exponential trend. It increases investment during undervaluation and reduces it during overvaluation, aiming to outperform buy-and-hold through adaptive allocation and then we use Monte Carlo simulations for future performance assessment.

3.1 Basic Principles

The weighted investment method incorporates the valuation factor defined in Equation 4. Since it is not possible to determine the optimal investment timing with certainty, the strategy is based on the principle that allocating capital earlier allows more time for compound interest to take effect, which tends to increase long-term returns [6]. One thing to note is that this principle becomes less relevant during periods of unsustainable market overvaluation (e.g., the Dot-com Bubble) and more relevant during undervaluation [6].

Following these principles, the monthly investment amount is dynamically adjusted using the relative difference between the actual value of the S&P 500 and its exponential trend. The allocation is calculated using the formula:

$$\text{Allocation} = \beta \cdot (1 - \alpha \cdot \text{difference}) \tag{5}$$

Where α is a sensitivity constant that determines how strongly the allocation reacts to deviations from the exponential trend and β represents the base monthly investment amount under neutral market conditions.

This model increases the investment when the index is below the expected exponential trend (indicating potential undervaluation) and reduces it when the index is significantly above the trend (indicating potential overvaluation). It offers a systematic alternative to discretionary market timing while maintaining exposure to long-term growth.

3.2 Practical Considerations and Allocation Constraints

A notable limitation of this model arises from its disregard for real-world tax implications and transaction costs. In many jurisdictions, frequent asset sales or rebalancing can trigger complex tax reporting requirements and significant costs, particularly if assets are sold before a minimum holding period. Since a comprehensive analysis of tax laws across different countries is beyond the scope of this study, a simplifying assumption is adopted: assets are never sold. This rule ensures the model avoids generating sell signals, maintaining compliance with the "buy-only" philosophy and reducing administrative burden.

Additionally, to enhance flexibility and reflect individual investment preferences or constraints, the model allows for an optional monthly investment cap. Each investor may define a maximum allocation amount per month, beyond which no additional capital is deployed, regardless of the valuation signal.

Incorporating these constraints, the final form of the weighted investment allocation is defined as follows:

$$\beta'(t) = \min(\beta_{\max}, \max(0, \beta(1 - \alpha\Delta\Phi))) \tag{6}$$

Here, β_{max} represents the investor-defined monthly allocation ceiling. This piecewise function guarantees that the allocation remains non-negative and bounded, aligning the theoretical model with practical investment constraints.

3.3 Backtesting

To assess the effectiveness of the dynamic allocation strategy, backtesting was conducted using a sensitivity coefficient of $\alpha = 2.5$, without imposing an upper limit on the monthly allocation. While the absence of such a cap allows for greater flexibility in investment decisions, it also introduces a potential constraint, as will be discussed in the subsequent sections.

Assuming a fixed monthly contribution of $\beta = 500€$, the model was applied retrospectively over the past 40 years, starting in 1985. The results, shown in Fig. 5, demonstrate that the weighted investment strategy achieved a higher final portfolio value compared to the traditional buy and hold approach. Specifically, while the buy and hold strategy yielded a final portfolio value of approximately 1.7 million euros, the weighted investment method reached a final value of approximately 2.0 million euros, representing an increase of 17%.

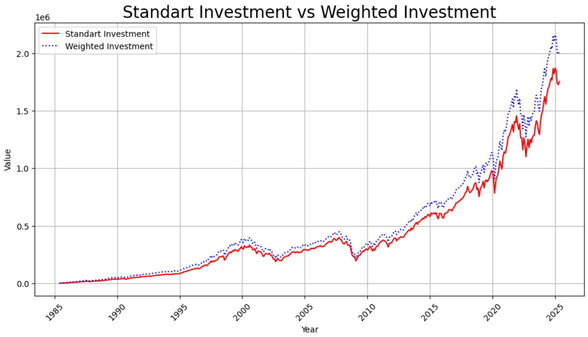

Figure 5: Buy and Hold vs. Weighted Investment – Portfolio Value

It is important to note, however, that the two strategies allocated different amounts of capital each month. Despite this, the difference in allocation amounts accounted for only 0.22%, a minimal variance that is unlikely to significantly influence the overall results.

Fig. 6 highlights how the allocation levels varied over time. This chart clearly shows how the model's adaptive nature responds to market phases, adjusting investment based on the deviations from the exponential trend line. In times of market overvaluation, the model reduced its allocation, while in periods of undervaluation, it increased the investment, reflecting the dynamic approach at play.

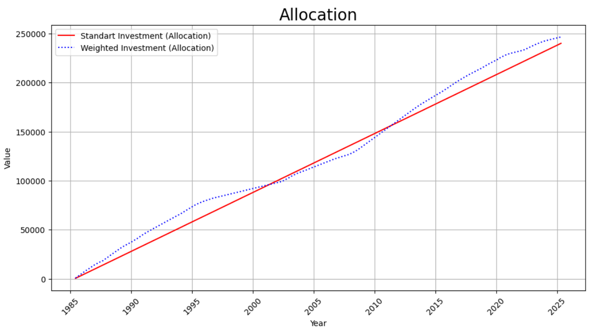

Figure 6: Monthly Allocation Over Time: Buy and Hold vs. Weighted Investment

3.4 Monte Carlo Simulations

Having conducted preliminary backtesting analysis and confirmed the proposed method's efficacy in historical context, we now examine its potential future performance. While testing with current market conditions might seem plausible, this approach would contradict the principles established in subsection 3.1, as the timeframe between analysis and results would be insufficient for robust long-term evaluation. Investment literature consistently recommends approximately 20-year periods for proper long-term assessment.

To address this limitation, we employ Monte Carlo simulations to explore potential S&P 500 price movements across various market scenarios. This approach provides a probabilistic performance assessment of the weighted investment strategy under different future conditions while respecting the necessary long-term perspective.

The Monte Carlo simulation [7] is employed to evaluate the future performance of the weighted investment strategy by simulating multiple potential future paths of the S&P 500 index. This technique allows us to capture the inherent uncertainty and variability of financial markets, providing a probabilistic distribution of possible outcomes. By simulating a large number of potential scenarios, Monte Carlo methods facilitate a deeper understanding of the range of potential future market conditions, thereby offering valuable insights for investment strategies.

Logarithmic Returns To model the evolution of asset prices, we utilize logarithmic returns, defined as the natural logarithm of the ratio between consecutive prices. Logarithmic returns are advantageous because they effectively capture the compounded nature of asset price changes and provide a better representation of returns over time. Specifically, the logarithmic returns are calculated as follows:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

where P_t is the price at time t, and r_t represents the return over a given time period. This transformation allows us to model the return distribution more efficiently while ensuring consistency with the continuous compounding process.

Estimation of Drift and Volatility Two key parameters govern the stochastic behavior of asset prices: the expected return (μ) and the volatility (σ). The expected return μ represents the average rate of return of the asset in continuous time, derived from the Compound Annual Growth Rate (CAGR):

$$\mu = \ln\left(\frac{P_{\text{final}}}{P_{\text{initial}}}\right)^{\frac{1}{T}} = \frac{\ln\left(\frac{P_{\text{final}}}{P_{\text{initial}}}\right)}{T},$$

where P_{final} and P_{initial} are the final and initial prices of the asset, and T is the investment period in years.

Volatility, defined as the standard deviation of the logarithmic returns,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2},$$

quantifies the degree of uncertainty or risk associated with price fluctuations.

In the context of Geometric Brownian Motion (GBM), the drift term used in simulations is not μ directly, but rather adjusted to:

$$\text{drift} = \mu - \frac{1}{2}\sigma^2,$$

which reflects the average growth rate of the log-price after accounting for variance.

In this analysis, μ is estimated from historical S&P 500 data, and σ is computed from the standard deviation of monthly log-returns.

Stochastic Process: Random Walk Model The random walk model is employed to simulate the future price trajectories of the S&P 500 index. At each time step, the price is updated according to a stochastic differential equation, capturing both the drift and the random fluctuations in the market. Specifically, the price update follows the equation:

$$P_{t+1} = P_t \exp\left(\mu - \frac{1}{2}\sigma^2 + \sigma Z_t\right)$$

where Z_t is a standard normal random variable, representing the random shock at each step, and P_t is the price at time t. The term $\mu - \frac{1}{2}\sigma^2$ adjusts the drift for the volatility, ensuring the model correctly reflects the risk-adjusted return over time.

Simulation of Price Paths The Monte Carlo simulation involves iterating the random walk model over multiple time steps, from the present to a defined future horizon. For each simulation, a random shock is drawn from a normal distribution, and the price is updated accordingly. We perform 10,000 simulations to generate a broad spectrum of potential outcomes, ensuring that the range of possible future market behaviors is adequately captured. This stochastic process allows us to explore the variability in the S&P 500 index's future performance.

Portfolio Simulation For each simulated price path, we simulate the evolution of a portfolio that invests a fixed amount of money each month. The number of units of the S&P 500 held in the portfolio is dynamically adjusted according to the changing price, based on the weighted investment strategy. This approach allows us to track the portfolio's value across all simulations and evaluate its potential performance under various market conditions. By simulating the portfolio across 10,000 price paths, we can assess the risk-return profile of the investment strategy in a probabilistic manner.

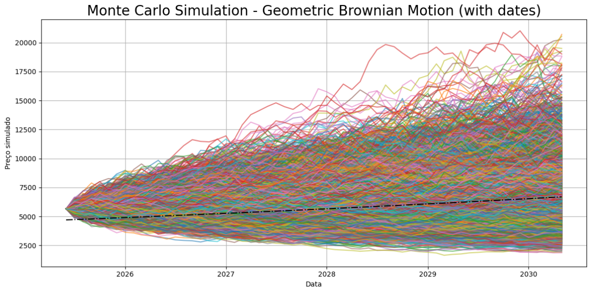

Figure 7: Monte Carlo Simulations

4 Empirical Study and Results

Analyzing the results obtained from the Monte Carlo simulations, we observe the following histogram:

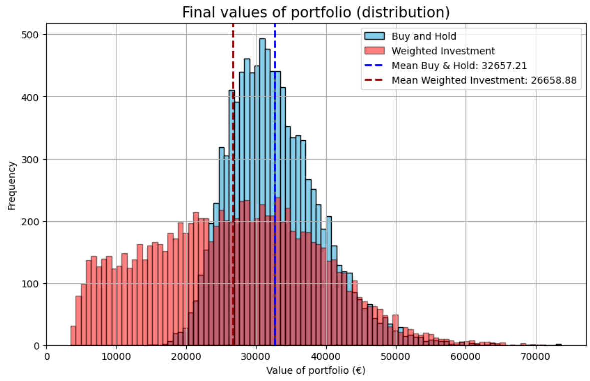

Figure 8: Final Portfolio Comparison Histogram

At first glance, the outcomes appear discouraging for those aiming to invest dynamically using the weighted investment strategy. However, upon critical reflection, it becomes clear that this histogram tells only part of the story. A careful reader will recognize that comparing the final portfolio values directly is misleading, since the capital allocated in the standard investment approach is fixed, whereas the weighted investment involves dynamic allocation, which may bias the results either positively or negatively.

To gain deeper insight, we examine the histogram of total allocations under the weighted investment strategy:

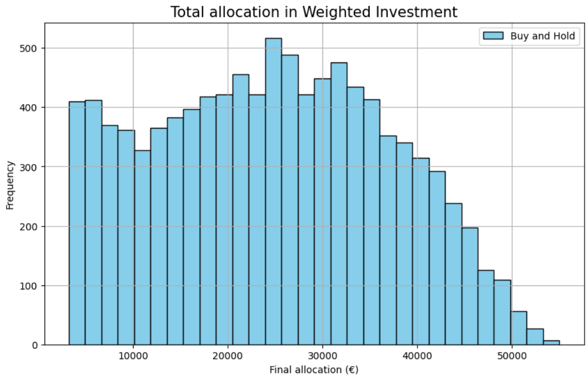

Figure 9: Total Allocation in Weighted Investment

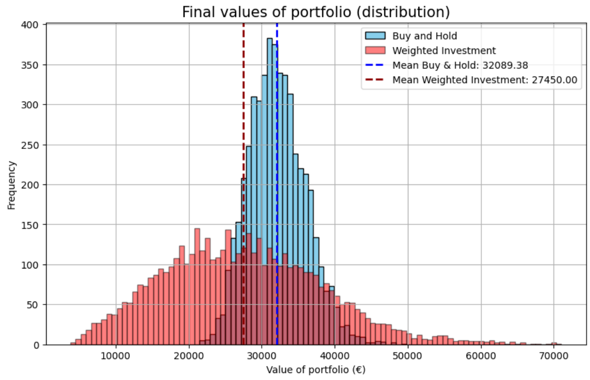

Figure 11: Final Portfolio Comparison Histogram

This discrepancy between values underscores the need for a more robust metric to compare performance. Therefore, we analyze the Return on Investment (ROI) for each method.

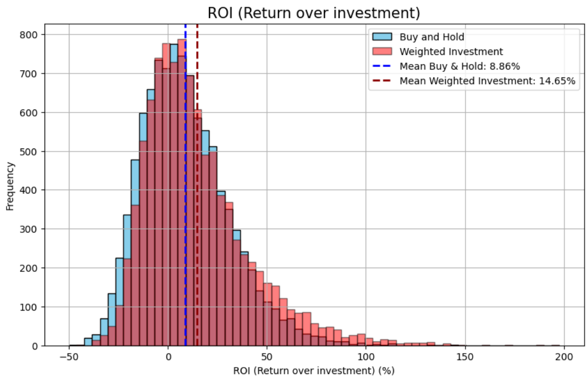

Figure 10: ROI Comparison Histogram

We can already observe a higher ROI with the weighted investment method, as initially anticipated. However, in extreme bull or bear market scenarios, it is possible to experience negative ROI with this approach, though for entirely different reasons. In the former case, the market is persistently overvalued, leading to minimal capital allocation; if the market then drops below a certain threshold, most of the invested funds become exposed during the decline, potentially resulting in a negative ROI. In the latter case, the reason is more straightforward: the market remains in a prolonged downturn, causing a continuous depreciation of the held assets.

Improvements We observe that the most extreme scenarios generated by the Brownian motion can adversely affect our method, as they lead to excessively aggressive allocations. Given that such extreme cases are highly unlikely in a real financial market, we focus our analysis on the interquartile range — specifically, the results between the 25th and 75th percentiles. By repeating the process within this narrower range, we obtain the following outcome.

As shown in Figure 11, even after excluding extreme scenarios, results improved little. While ROI still exceeds buy-and-hold, the portfolio fails to surpass nominal value. The reason is simple: initial conditions. As of May 1, 2025, the market is 20% overvalued by our method, weakening the long-term investment's impact, despite it holding the greatest weight, as explained in subsection 3.1.

For this very reason, we tested the same simulation assuming the current date was January 1, 2001, a period when the market was close to its expected value. The results were as follows:

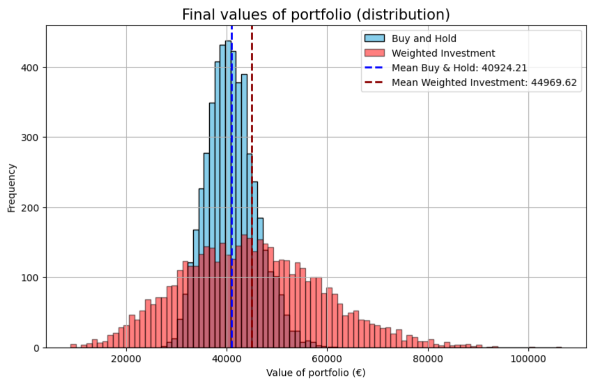

Figure 12: Final Portfolio Comparison Histogram - 2001

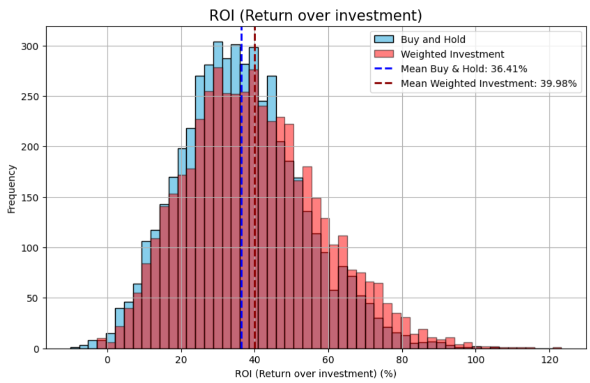

Figure 13: ROI - 2001

These most recent results suggest that the weighted investment strategy may be a superior option for investors who feel they are missing market buying opportunities. By dynamically allocating more capital during downturns, this approach enables investors to potentially amplify gains when the market recovers. However, it is important to note that the strategy is highly sensitive to the initial market conditions at the time of investment. This dependency introduces a structural vulnerability that may limit its robustness across different market cycles. A potential avenue for improvement could involve the development of a hybrid investment strategy that combines the strengths of both dynamic and static allocation models. This is one of the directions proposed for future research in Section 5.

5 Conclusion, limitations and future work

In this paper, we explored a dynamic allocation strategy for the S&P 500 that operationalizes value investing principles [6]. This approach systematically adjusts capital deployment based on quantitative valuation metrics, demonstrating improved risk-adjusted returns in backtesting while mitigating common behavioral pitfalls [8]. Key advantages include: (i) **rule-based implementation**; (ii) **downside protection**; and (iii) **compatibility with dollar-cost averaging**.

However, practical limitations persist, particularly regarding capital availability and trade-offs related to compounding, as noted in [9]. Notably, we found that this strategy challenges Graham's well-known principle that "time in the market beats timing the market" [6], by systematically reducing early-stage investments during periods of overvaluation, leading to a form of **compounding disruption**. Moreover, while [9] emphasizes the importance of consistent, regular investing, our strategy requires disproportionately larger contributions during market downturns, precisely when personal liquidity is often limited. This introduces a behavioral hurdle highlighted by [8], which we refer to as the **capital availability paradox**. Furthermore, the monthly adjustments increase **execution complexity** and demand the disciplined mindset described by [10]. However, this level of discipline often exceeds what most individual investors can sustain consistently without the aid of automated tools, such as those enabled by [2].

Future automation and nonlinear refinements could enhance the dynamic allocation principle while preserving the core valuation discipline. We would like to explicitly suggest: (i) follow [6]'s margin of safety principle by reducing exposure during extreme overvaluation; (ii) operationalize [8]'s "room for error" concept through dynamic allocation; (iii) provide systematic rules that counter [10]'s identified "scarcity mindset" during downturns.

Future Research Directions This study opens several promising avenues for further development.

First, the linear allocation mechanism in Eq. 5 could be refined by incorporating non-linear responses to valuation signals. Functions such as sigmoid or threshold-based activations may better reflect behavioral finance insights from [8], responding more strongly to extreme market conditions.

A practical extension would involve building an open-source platform integrating data sources like [2] with brokerage APIs to automate the proposed strategy. This could support both educational and applied use, enabling real-time dynamic allocation.

Another avenue is analyzing the tradeoff between early investing and waiting for better valuations. This depends on personal financial factors such as savings patterns [9], income uncertainty, and macroeconomic shifts, which future work could help model more precisely.

Finally, applying this framework to other asset classes like global equities, real estate, or cryptocurrencies could test its robustness and lead to asset-specific adaptations.

References

[1] W. E. Buffett, “The superinvestors of graham-and-doddsville,” *Hermes*, 1984. [Online]. Available: https://business.columbia.edu/cgi-finance/chazen-global-insights/superinvestors-graham-and-doddsville

[2] R. Aroussi, “yfinance: Market data downloader library,” https://pypi.org/project/yfinance/, 2023, version 0.2.18.

[3] Brains.dev. (2022) Hands-on linear regression with python code. Brains.dev. [Online]. Available: https://brains.dev/2022/pratica-regressao-linear-com-codigo-em-python/

[4] D. Hackers. (2023) Implementing simple linear regression in python.

[5] D. Team. (2023) Linear regression in python: Tutorial. DataCamp. [Online]. Available: https://www.datacamp.com/pt/tutorial/linear-regression-in-python

[6] B. Graham, *The Intelligent Investor*, revised ed. Harper Business Essentials, 2003, originally published in 1949.

[7] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, ser. Applications of Mathematics. New York: Springer, 2004.

[8] M. Housel, *The Psychology of Money*. Harriman House, 2020.

[9] G. S. Clason, *The Richest Man in Babylon*, updated ed. Penguin, 2010, originally published in 1926.

[10] T. H. Eker, *Secrets of the Millionaire Mind*. HarperCollins, 2005.