**ERASMUS UNIVERSITY ROTTERDAM**
**ERASMUS SCHOOL OF ECONOMICS**
**Bachelor Thesis Economics & Business**
**Specialization: Financial Economics**

# Smooth Operators
## The Roles of Volatility Forecasting and Scaling in Avoiding Momentum Crashes

**Author:** Stoppani Andrea
**Student number:** 602751
**Thesis supervisor:** Prof. Marshall (Xiaoyin) Ma
**Second reader:**
**Finish date:** Friday 29th June 2024

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second reader, Erasmus School of Economics or Erasmus University Rotterdam.

## ABSTRACT

This paper studies the impact of volatility forecasting methods on the performance of dynamically scaled momentum strategies in the U.S. equities market. The volatility forecasting methods used include: historical mean variance, Monte Carlo simulations, GJR-GARCH models, and machine learning algorithms. The enhanced strategies significantly outperform the traditional momentum strategy (MOM) and the market. Using mean historical variance yields the highest mean returns and Sharpe ratio, making the method most attractive for risk-seeking investors. The model based on GJR-GARCH forecasts excels in risk management; this portfolio has the lowest maximum drawdowns and highest positive skewness, making it the best for risk-averse investors. The machine learning approach has the best performance in recent years, especially during financial crises, demonstrating resilience and adaptability. These findings highlight the importance of selecting appropriate volatility forecasting methods to enhance momentum strategies' performance.



**Keywords:** Volatility Scaling, Volatility Forecasting, Momentum, Monte Carlo Approach, GJR-GARCH, Machine Learning

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1 Introduction

Lawrence Summers, former secretary of treasury of the United States, is widely attributed with saying, "Most investors want to do today what they should have done yesterday" (Summers, n.d.). While there is no doubt that it would be better to have acted yesterday, acting today might still prove to be enough. Momentum strategies are a prime example of this as they consist in going long on winner stocks, and short on losers, and tend to be significantly profitable. In the period between January 1930 and December 2017, a long-short momentum factor portfolio of U.S. equities would have generated a raw average return of 0.60% per month, with a Fama French 3 Factor model alpha of 0.88% (Hanauer and Windmuller, 2022). For this reason, since their introduction (Jegadeesh and Titman, 1993) momentum strategies have been frequently used in investing, and are still in use today, especially on securities with higher volatility such as cryptocurrencies (Bloomberg, 2024). However, despite its profitability and widespread use, momentum has its downsides. Momentum leads to especially large dropdowns, in particular during market reversals as documented by Asem and Tian (2010). Thus, identifying effective solutions to minimize these drawdowns has been a hot topic in research ever since these strategies were introduced.

Previous papers which study methods to mitigate the effects of momentum crashes suggest various solutions to the problem, such as implementing a 10% stop-loss order (Han et al., 2016), and reversing the strategy after a significant local market crash (Dobrynskaya, 2019). Other authors explore volatility-scaling strategies which consist in determining a position's size based on its volatility, to contrast the leverage effect. This results in larger (smaller) positions when volatility is lower (higher). Three volatility-scaling strategies have been proposed and studied: constant volatility-scaled momentum (Barroso and Santa-Clara, 2015), constant semi-volatility-scaled (downside risk) momentum to distinguish between up and down movements (Wang and Yan, 2021), and dynamic-scaled momentum which also considers the estimated return (Daniel and Moskowitz, 2016). Hanauer and Windmuller (2022) compare these strategies across equity markets of 49 countries. They find that volatility-scaling is successful in reducing momentum crashes and yields higher Sharpe ratios. To predict volatility, the authors use the average squared daily returns of the past 6 months following Barroso and Santa-Clara (2015). This estimation method has the advantage of simplicity; however it ignores any current market conditions that might affect future volatility, and it does not account for changes in volatility patterns over time.

This paper explores how different volatility estimation methods affect the returns and overall performance of the dynamically scaled momentum strategy. To estimate the models, I will use the following additional volatility estimation methods: a Monte Carlo simulation following the Geometric Brownian Motion (GBM) model, a GJR-GARCH model, and a machine learning algorithm based on the random forest method. The Monte Carlo approach with GBM has the advantage of being forward-looking, thus not having to rely on past data to make a forecast. On the other hand, the GJR-GARCH model considers the time-varying nature of volatility while accounting for its asymmetric shocks. Finally, the machine-learning model has the advantage of not requiring any initial assumptions. I will also estimate volatility following Barroso and Santa-Clara (2015) and use this value for the reference portfolio. It is expected that these enhanced models, with improved volatility estimation processes, will have better performances due to their more accurate volatility forecasts. Thus, in this paper I will study:

*How does the volatility estimation process affect returns and overall performance of enhanced momentum strategies in the U.S. equities market?*

To compare the strategies, I will use data on the U.S. stock market from January 1st, 1962, to June 1st, 2024. The main unit of this analysis will be daily returns, which will be computed based on the daily close value for each security, retrieved from Yahoo Finance. I will then compare portfolio performances based on different metrics. Mean returns will be used to evaluate overall performance, the latter will have the relevant t-statistic to test their significance. Additionally, I will compare Sharpe ratios to examine the risk weighted returns, and different measures of portfolio risk including volatility, skewness, kurtosis, and maximum drawdown. Finally, I will estimate the maximum transaction cost that renders profits insignificant at a specific level of $\alpha$ following Grundy and Martin (2001), and Barroso and Santa-Clara (2015). Graphs of the portfolios' performances, reported as cumulative returns, will also be provided. This will help better assess how the portfolio performances are affected by the different volatility forecasts over time. I will then regress the portfolios based on the Fama French 3 Factor model (Fama & French, 1993) to observe if the strategies yield a significant $\alpha$, which would be interpreted as unexplained return based on the chosen risk factors. Finally, I will perform pairwise tests with t-tests and the sign test, to verify whether any strategy significantly outperforms the other. The sign test will be included to account for the potential non-normality of portfolio returns.

In this analysis, the more advanced and complete volatility estimation methods are expected to improve the effect of volatility scaling on the momentum strategies, as opposed to using the mean squared returns of the past 6 months as a forecasted volatility proxy. Additionally, I expect this difference to be particularly noted in the reduced skewness, kurtosis, and maximum drawdown as a more precise estimate of risk is vital in pre-emptively mitigating its downside effects. This research will add to previous research on the topic, as it will offer a more comprehensive analysis on the most used and currently trending volatility forecasting tools and on their applications in portfolio risk management, more specifically for momentum portfolios. Additionally, these results will provide further insight on momentum strategies, and on how they can be adapted and improved to become feasible investment mechanisms for any type of investor. However, I expect better and improved volatility forecasting models to arise as volatility forecasting methods become more sophisticated and accurate. Additionally, as different approaches such as those of Han et al. (2016), Dobrynskaya (2019), and others are implemented and improved, better methods may arise. Finally, developments in volatility forecasting, and data processing might better explain the nature of momentum crashes by approaching the problem from a different perspective.

The empirical results of this study do not allow us to identify a comprehensively better strategy out of the enhanced momentum strategies. Overall, the dynamically scaled momentum (dMOM) portfolios consistently outperform the basic momentum strategy and the market, in the sample. When using the historical mean variance as a forecasting mechanism for volatility scaling, the portfolio that yields the highest mean returns and Sharpe ratio is obtained. On the other hand, using GJR-GARCH models to forecast volatility for volatility scaling, yields the best performance as goes for risk-management, due to the portfolio's low maximum drawdown, and its high and positive skewness. Finally, when considering the most recent period in the sample, between June 2004 and June 2024, the dMOM model, based on machine learning algorithm forecasts yields the best performance based on cumulative returns. These results highlight the fact that more advanced volatility forecasting improves risk management capabilities, however the correct model should be chosen based on an investor's risk appetite.

The paper will be structured as follows. To start, the literature review will consider the most relevant existing studies on the topic, and briefly summarize their findings. This will lead to the formation of the primary research question along with the sub-questions. Then, the data and methodology sections will describe the process of deriving portfolio returns, and the methods which will be used in the statistical analysis. Subsequently, the results of such analysis will be presented. Finally, the conclusion will include a discussion, address limitations, and offer suggestions for further research based on the analysis.

# CHAPTER 2 Theoretical Framework

## 2.1 Volatility Forecasting

In this section the topics of volatility forecasting, portfolio performance evaluation, and momentum strategies will be discussed. The discussion will be based on previous papers and includes a historical review of these topics.

In this paper, I will use the outputs of different volatility forecasting methods to scale a momentum portfolio, based on the dynamic-scaled momentum strategy (Daniel and Moskowitz, 2016). Thus, before proceeding, it is crucial to properly define volatility. Volatility is not a concept strictly related to finance and has slightly different interpretations based on the field. In finance, volatility is the rate at which the price of a security, or the returns of a portfolio or stock increase or decrease, given a time horizon (cf. Forbes, 2023). In more practical terms, volatility is a measure of expected market fluctuations, and a proxy for market risk. Two main categories of volatility exist: implied volatility which is the value of volatility factored in the security's current option price, and realized volatility which is based on historical performance. However, it is not clear if one measure outperforms the other, Christensen and Prabhala (1998) find that both methods are equivalent when forecasting realized volatility. Thus, in this paper, the realized volatilities calculated using log-returns will be used as a volatility proxy for the calculations.

The root of modern quantitative finance and the use of metrics such as volatility in investment decisions, is usually attributed to Louis Bachelier who introduced the concept of Brownian motion (1900). Building on the work of Bachelier, Metropolis and Ulman (1940), introduced the "Monte Carlo Method", a computational technique based on random sampling to make projections. Volatility finally became a mainstream topic in finance, when Benoit Mandelbrot (1960), highlighted the presence of large, unpredictable market moves. This paper influenced the understanding of volatility, stimulating researchers to find better ways to forecast these movements. One of the most disruptive discoveries for financial markets was when Black and Scholes (1973) and Merton (1973), revolutionized the pricing of options and incorporated volatility as a crucial factor. This model (Black-Scholes formula) allows for the calculation of implied volatility and is still in use today. Similarly, Engle (1982) disrupted the industry by introducing a whole new category of models for volatility forecasting, the Autoregressive Conditional Heteroskedasticity (ARCH) models. ARCH models are used to forecast realized volatility based on its lagged values. Currently trending volatility forecasting techniques include improved and more complex versions of Engle's ARCH model, deep learning models, and long-memory models.

When considering the volatility forecasting techniques used in this paper, evidence can be found of their successful application in finance. Glasserman (2004) discusses the effectiveness of the Monte Carlo method in modelling the stochastic nature of financial markets. The author suggests that with an appropriate number of simulations, one can get a clearer idea of the likelihood of extreme events and potential risks. Nugroho et al. (2019), find that the GJR-GARCH model provides the best fitting results compared to any other GARCH family model. This is due to the fact that the model effectively captures the leverage effect and accounts for the asymmetry between positive and negative shocks. Finally, the empirical findings of Diane and Brijlal (2024), show that Random Forest algorithms outperform Artificial Neural Networks in volatility forecasting in the South African stock market. However, no research has been done comparing these methods, especially when applied to volatility scaling.

### 2.2 Portfolio Performance Evaluation

Having discussed the input of interest, the paper will proceed by mentioning how the process of evaluating portfolios has evolved over time. To start, there is no universally accepted framework to evaluate a portfolio's performance, as different methods should be used in different situations. The classic approach to evaluating a portfolio's performance is limited to evaluating the portfolio's returns, cumulative returns which help visualize the portfolio's performance over time, and the Sharpe ratio which provides a measure of risk-weighted returns. These values are then compared to a benchmark, usually the market. While these methods provide straight-forward information on a portfolio's performance, they have significant limitations. Evaluating the Fama French 3 Factor model $\alpha_{FF3,s}$ of a portfolio is arguably a better measure than using mean returns when comparing performance with the market. By considering the $\alpha_{FF3,s}$, one controls for factors such as firm size and the book-to-market factor which would not be possible when basing the analysis on mean returns (Mahrfor, 2015). Additionally, Kolari, Liu and Pynnönen (2024) recommend including measures of potential drawdowns such as kurtosis, skewness, and maximum in-sample drawdown. When evaluating a portfolio, this is especially for portfolios including short positions, as large, unexpected drawdowns might lead to margin calls. In this paper I will not use a single or universal measure of performance, as such a method does not exist. However, I will use different methods to evaluate different aspects of a portfolio, such as its risk and return in order to have the most complete overview of the portfolio's performance.

Since the start of investing, finding the best measure or method to evaluate a portfolio and compare it to other options has been a central topic in finance. Not being able to distinguish which portfolio performs best, or not considering crucial factors such as expected risk will most likely lead to wrongful decisions and potentially huge losses. One of the first, and most influential works on the topic is "Mutual fund performance, in which Sharpe introduced the "Reward-to-Variability ratio" currently better known as the Sharpe ratio (1966). This measure allows to calculate the risk adjusted return of a portfolio and became a steppingstone for future research. Later, Jensen introduced the Jensen alpha, based on the CAPM model (1968). This method allows to identify whether the strategy yields return in excess of the market. Many extended versions of the model started being used in the following years, mainly based on Fama French models. More recently, Daniel, Sornette, and Wohrman, published an influential paper focusing on the issue of look-ahead bias in back-testing investment strategies (2008). This study considers portfolio evaluation from a whole different perspective and provides insight on how to deal with this bias. More recently, models that attempt to integrate behavioral finance insights and ESG metrics are gaining traction, due to the increasing ability to process large quantities of data and the fact that investors are becoming more conscious of environmental, social, and governance factors (Khan, Serafeim, and Yoon, 2016).

### 2.3 Momentum

Finally, I can proceed with discussing the strategy I will study in this paper. The general definition of momentum interprets it as a property of a moving body which determines the amount of time required to stop its motion, when a constant force is applied (cf. Merriam-Webster online, 2024). Thus, the concept of momentum is originally related to physics, however it is used also in other contexts. In finance, a momentum strategy is a type of investment strategy based on the assumption that stocks, like objects, have inherent momentum. This implies that once a stock is moving upwards or downwards (in value) this movement will continue for a certain period. More specifically, momentum strategies consist in going long (short) stocks that have had a positive (negative) trend in the past period. Overall, momentum strategies have been proven to have positive $\alpha$, and to return better risk adjusted returns than the market (Jegadeesh and Titman, 1993). The classic momentum strategy is a quantitative type of investment strategy which implies that it is a strategy which does not consider the business model, management team, and other more qualitative factors when making an investment decision.

According to various sources, momentum investing dates to as early as the Victorian Era (Chabot, Ghysels, and Jagannathan, 2014), however Richard Driehaus is often referred to as the father of momentum investing as he is accredited with popularizing the strategy in the 80s (Larsen and Larsen, 2018). One of the first academic articles providing empirical evidence on momentum, written by Levy, in the late 60s (1967) proving that stock that performed well in the past continue performing well. However, the article did not gain much traction, unlike “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency” (Jegadeesh and Titman, 1993) which is the seminal paper on the topic. In 1997, Carhart published a particularly influential paper, introducing an extended version of the Fama-French three-factor model which included a momentum factor. This helped prove that momentum is an economically and statistically significant determinant of portfolio returns, emphasizing its importance in portfolio formation. More recently, alongside others, Blitz, David, and Pim van Vliet (2008), and Asness, Moskowitz, and Pedersen (2013) proved the efficiency and profitability of momentum strategies over various asset classes and markets worldwide. However, even if the classic momentum strategy is widely regarded as a profitable investing strategy, serious concerns have risen due to its high maximum drawdowns which make the strategy unfeasible or disregarded by more risk-averse investors (Barroso and Santa-Clara, 2015). Thus, finding ways to mitigate these risks has been a hot topic in the past decade. Han et al. (2016) have suggested the simple but effective solution of implementing a 10% stop-loss order which more than halves the maximum monthly loss and doubles the strategy's Sharpe ratio. Dobrynskaya (2019) suggests reversing the strategy after a significant local market crash, to contrast the effect of market reversals and turn them into opportunities to profit.

A different approach consists in utilizing volatility scaling which is the process of increasing or decreasing the portfolio's size based on the forecasted volatility. The purpose of volatility scaling is to contrast the leverage effect, which is an empirically proven stylized fact of finance (Wang and Mykland, 2014). Generally, the leverage effect consists in the notion that when volatility is high returns are low. Additionally, one must mention that scaling a momentum portfolio is feasible, as it is a long-short strategy. Which implies that the standard momentum portfolio can be theoretically scaled indefinitely due to its self-financing nature. Barroso and Santa-Clara (2015) are the first to come up with a volatility scaled momentum strategy. More specifically the constant volatility-scaled momentum, which scales the portfolio based on a ratio of a volatility target, divided by the forecasted volatility for the next period. Wang and Yan (2021) introduced the semi-volatility-scaled momentum, making use of Roy's (1952) semi-volatility intuition, to distinguish between up and down movements, and thus hedge against the downside risk. Finally, Daniel and Moskowitz (2016) suggest using a dynamic-scaled momentum, which has the main advantage of allowing the portfolio weight to be negative in case the forecasted return is negative. The scaling weight is calculated based on an adapted version of the Markowitz model (1952) and aims to maximize the risk-weighted return. The biggest advantage of this solution is its ability to adapt the portfolio to market conditions similarly to the method proposed by Dobrynskaya (2019), alongside hedging for the risk of high market volatility.

In their work on enhanced momentum strategies, Hanauer and Windmuller (2022) compare the strategies of Barroso and Santa-Clara (2015), Wang and Yan (2021), and Daniel and Moskowitz (2016) across equity markets of different countries. In their analysis, the authors use the method used in Barroso and Santa-Clara for volatility forecasting, which consists in calculating the average daily volatility over the past 6 months and turning it into monthly terms. This choice is made for comparative reasons, to better evaluate the performance of the different strategies. The authors conclude that no strategy significantly outperforms the other, however the dynamic-scaled momentum strategy has reduced skewness, kurtosis, and maximum drawdown. This indicates an overall mitigated level of risk compared to the other strategies. Mindful of these results, in this paper I will be basing my analysis on the dynamic-scaled momentum strategy following Daniel and Moskowitz (2016). As volatility scaling for momentum strategies is a niche topic, not much research has been done on which volatility forecasting approach better suits this technique.

In this study, I replicate the work of Daniel and Moskowitz (2016) due to their method yielding the best performing portfolio in the empirical comparison performed by Hanauer and Windmuller (2022). I test different volatility forecasting techniques and their effect on the dynamic-scaled momentum strategy. These methods include calculating the mean volatility over the past 6 months following Barroso and Santa-Clara (2015), a Monte Carlo simulation, a GJR-GARCH model, and a Machine Learning algorithm integrating insights from the other models. Compared to the base model (mean volatility), the Monte Carlo approach has the advantage of being forward looking, the GJR-GARCH model has the advantage of considering the time series structure of data, and the Machine Learning model has the advantage of not making any initial assumptions on distributions or data structures. In the study I will use data from the US stock market, where the momentum strategy has been empirically proven to work (Levi, 1960). As returns tend to be normally distributed, given the number of simulations is sufficiently large, the first hypothesis is:

**H1:** *The Monte Carlo approach does not improve the portfolio's performance when compared to the base (mean volatility) model.*

Due to the GJR-GARCH more detailed understanding of the data structures, the second hypothesis is as follows:

**H2:** *The GJR-GARCH volatility estimation technique improves the portfolio's performance when compared to the base (mean volatility) model.*

Finally, I formulate the last hypothesis. As the machine learning model has the most complete input data, and makes no initial, strict assumptions, I expect that:

**H3:** *The Machine Learning method brings the greatest improvement to the portfolio, compared to the other methods.*

More generally, in this study, I will assess the portfolio improvement brought by the different volatility forecasting methods. The process will not only assess the mean and cumulative return difference, but also compare the level of risk of the strategies.

# CHAPTER 3 Data and Methods

## 3.1 Raw Data

The data used in this study consists of daily close values for 9,990 stocks listed in the United States. Due to the dataset not being clean, multiple filters were applied. Penny stocks were removed from the data, defined as stocks having a price per stock under 5 USD on June 1st, 2024. All missing values, likely generated due to differences in trading days between exchanges, were forward filled to match the close value of the next trading day. Additionally, all stocks that had a day-to-day return exceeding 50% were removed from the sample, as such extreme returns are outliers. French Schwert and Stambaugh (1995) find that high returns (defined as 50% daily increases) are rare and typically associated with extraordinary events or speculative bubbles. These filters remove stocks which are more likely to be subject to market manipulation due to low liquidity and smaller market cap, and overall reduce noise in the data. However, as a consequence the overall sample volatility is significantly reduced. It must be mentioned that filtering for extreme day-to-day gains is a particularly conservative approach; however, by doing so, all unrealistic observations are excluded, and momentum signals are better isolated to avoid false positives and better assess the strategy's performance. After the data cleaning process, the sample is reduced to 6,900 securities. The data constitutes of daily adjusted close values starting in January 1962 and ending in June 2024 and is obtained through Yahoo Finance's python API. This period was chosen, as January 1st, 1962, is the first, available data point on the platform. Yahoo Finance is a financial platform that offers real-time and historical market data, financial news, and investment tools. Its data is retrieved from a variety of sources, including financial exchanges, public companies, and third-party data providers. The "yfinance" Python API facilitates access to this database. The adjusted close price, is the closing price of the day, adjusted to reflect the effect of corporate actions, such as dividends, stock splits, and rights offerings. This figure is used instead of the standard daily close price, as it better reflects the true economic value of holding the stock over time. The adjusted close prices are expressed in USD. Additionally, daily and monthly data on the Fama-French factors is obtained from Kenneth French's website[1]. More specifically: the risk-free rate (RF), the market excess return (RMRF), the average return on the three small portfolios minus the average return on the three big portfolios (SMB), and the average return on the two value portfolios minus the average return on the two growth portfolios (HML). SMB and HML values are calculated using the following formulas:

$$SMB = \frac{1}{3}(Small\,Value + Small\,Neutral + Small\,Growth) - \frac{1}{3}(Big\,Value + Big\,Neutral + Big\,Growth) \quad (1)$$

[1] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research

$$HML = \frac{1}{2}\ (Small\ Value\ +\ Big\ Value) - \frac{1}{2}\ (Small\ Growth\ +\ Big\ Growth) \qquad (2)$$

The relevant values are calculated based on Fama and French's paper: "Common risk factors in the returns on stocks and bonds" (1993).

**Table 1: Descriptive Statistics.**

| | Mean | Stand. Dev. | Max. | Min. |
|---|---|---|---|---|
| Adj. Close | 3,026 | 533,958 | 210,426,160 | 9.940e-11 |
| N. of Obs. | 3,042 | 3,251 | 15,721 | 1 |

*Notes*: The table includes the summary statistics related to the adjusted close values of the securities in the sample, and the number of observations. The "Mean", "Stand. Dev.", "Max.", and "Min." represent respectively the mean, standard deviation, maximum, and minimum. All values are expressed in USD for Adj. Close, and in units for N. of Obs., and are rounded to the next dollar. "e-11" stands for $\cdot\ 10^{-11}$.

As can be seen in Table 1, the adjusted close values have a considerably wide range (between $9.940 \cdot 10^{-11}$ and 210,426,160) and vary a significant amount from the mean on average. This is due to the fact that the adjusted close values are normalized to account for stock splits to provide a linear measure of the company's performance. This inevitably leads to exorbitant maximum and minimum values. The maximum value was registered by Direxion Daily Semiconductor Bear 3X Shares (ticker symbol: SOXS). SOXS is a 3x levered ETF offering short exposure to the semi-conductor market. Due to the recent boom in the industry, the index experienced a substantial decline in value. Due to this, the stock underwent multiple reverse stock splits to maintain a reasonable stock price. On the other hand, the minimum value was registered by Burke & Herbert Financial Services Corp. (ticker symbol: BHRB). The stock had an outstanding performance since its IPO and had to announce multiple stock splits such as a $40 - 1$ split in October 2022. Having a wide range of prices is not a concern as once the close values are converted to returns, they will be uniformized. Relative performance is of interest for this study, and not its absolute value. Similarly, the number of observations per security is subject to large variations, as it can observed that the standard deviation is larger than the mean. This is also not a main concern for the analysis as it does not affect the formation of momentum strategies.

### 3.2 Portfolio Construction

In this section I will discuss the process of portfolio formation, starting from the standard momentum portfolio, to then discuss the calculation of the relevant scaling weights for the dynamic-scaled momentum strategies (dMOM).

Starting from the adjusted close values of individual stocks, I calculate the respective returns. In the analysis, I used the individual stock returns to estimate the final portfolio returns. This process requires multiple steps, the first of which is calculating the returns of a long-short momentum portfolio, based on Jegadeesh and Titman (1993). For this study, a long-short momentum strategy with a 12-month lookback period, a 1-month gap, and a 1-month holding period is being used as the base strategy (MOM). These parameter choices are in line with the recommendations of Jegadeesh and Titman (1993), and with the model used by Hanauer and Windmüller (2023) except for the lookback period, about which the authors are not specific. In other words, to estimate MOM returns, given my parameters of choice, one must evaluate the cumulative performance of assets over the previous 12 months and identify the tickers in the top and bottom decile. Then, after a 1 month (gap period), the strategy involves going long on the top-performing assets and going short the bottom-performing assets. Thus, a decision to trade today would be based on the assets' cumulative performance in the months $t-2$ to $t-13$ (this number accounts for the 1-month gap period). These positions are then held for 1 month before re-evaluating and rebalancing the portfolio based on the new 12-month performance data. Additionally, as a long-short portfolio is being considered, one must sum the daily (monthly) risk free rate to the daily (monthly) return. This is due to the fact that a long-short strategy requires the investor to have a margin account, on which the risk-free rate is earned. Having calculated the returns of MOM, it is then necessary to calculate the appropriate weights for the dynamic-scaled momentum strategy (dMOM).

The implementation of volatility scaling to enhance MOM, has the objective of managing its realized volatility. Barroso and Santa Clara (2015), have shown that for cross sectional momentum (MOM), realized volatility has a positive correlation with future volatility, and negative correlation with future returns. It must be noted that volatility scaling itself, including the standard dMOM, does not account for the positive autocorrelation of returns. By combining these insights, one can generate a momentum strategy with an overall improved performance. Additionally, one must consider that since long-short strategies are self-financing, the portfolio can be theoretically infinitely scaled positively or negatively (going long the past losers and short the past winners). The dMOM approach, based on Daniel and Moskowitz (2016) is applied as follows to calculate the appropriate portfolio weights ($W_{dMOM,t}$):

$$W_{dMOM,t} = \left(\frac{1}{2\lambda}\right) \cdot \frac{\hat{\mu}_t}{\hat{\sigma}_t^2} \quad (3)$$

In the above formula, $\hat{\sigma}_t^2 = \mathbb{E}_{t-1}[\sigma_t^2]$ is the forecasted respective conditional expected variance of MOM, $\hat{\mu}_t = \mathbb{E}_{t-1}[\mu_t]$ is the forecasted respective conditional expected return of MOM, while $\lambda$ is a static scalar scaling the volatility of dMOM to that of MOM. Both $\hat{\sigma}_t^2$ and $\hat{\mu}_t$ are estimated using out-of-sample approaches, to prevent incurring in look-ahead bias. The return of MOM ($\hat{\mu}_t$) is estimated based on the following regression, using 126 trading days (6 months) of data:

$$R_{MOM,t} = \gamma_0 + \gamma_{int} \cdot I_{Bear,t-1} \cdot \sigma_{RMRF,t-1}^2 + \epsilon_t \tag{4}$$

In the regression formula, $I_{Bear,t}$ is a bear-market indicator at time t, $\sigma_{RMRF,t}^2$ is the realized variance of RMRF over the past 126 trading days (6 months) at time t, $\gamma_0$ is the regression intercept, and $\gamma_{int}$ is the coeffiecient of the interaction effect between $I_{Bear,t}$ and $\sigma_{RMRF,t}^2$. More specifically, $I_{Bear,t}$ is a dummy variable taking a value of 1 if the cumulative past two-year market return is negative, and 0 otherwise. In their study Daniel and Moskowitz (2016) find that momentum returns are particularly low in periods of market stress (bear market and high market variance), and that the optionality associated with loser portfolios is significant only in bear markets. This yields a significant negative market exposure during periods of market reversal (from bear to bull). Based on these intuitions, to hedge against the risk of market reversals, the authors include $I_{Bear,t}$ and $\sigma_{RMRF,t}^2$, to forecast MOM returns for the purpose of volatility scaling. Additionally, Daniel and Moskowitz (2016) find that MOM returns are particularly poor during bear markets with high volatility, thus justifying the choice of the interaction effect. The expected return, in Formula 3 is defined as the fitted values from this regression.

To forecast the expected variance $\hat{\sigma}_t^2$, I used 4 different methods: the mean historic variance, a Monte Carlo simulation following the Geometric Brownian Motion (GBM) model, a GJR-GARCH model, and a machine learning algorithm based on the random forest method. To calculate the mean historic variance, following Barroso and Santa-Clara (2015), I used the following formula where $R_{MOM,d-j,t}^2$ is squared realized daily return of momentum.

$$\hat{\sigma}_{MOM,t}^2 = 21 \cdot \sum_{j=1}^{126} \frac{R_{MOM,d-j,t}^2}{126} \tag{5}$$

For the Monte Carlo simulation, I used the squared realized daily return of momentum returns $R_{MOM,d-j,t}^2$, as a volatility proxy. The simulation has a sample of 126 trading days (6 months), and runs 1,000 simulations based on the normal distribution, to make its forecasts. Similarly, the GJR-GARCH method has a window of 126 trading days (6 months) as a sample to make its forecasts. Finally, the machine learning method uses a random tree algorithm to make predictions, based on a training sample of 126 trading days (6 months). The model uses 100 decision trees and is set to the random seed “42”, in order to ensure reproducibility of results. A new model is trained for every prediction, based on the most recent out-of-sample data. The training dataset includes the forecasts of the other models (historical mean variance, Monte Carlo approach, and GJR-GARCH), the past MOM returns, the past squared MOM returns, and the rolling variance estimated using the previous 30 observations. While the target dataset uses square returns as variance proxies.

Finally, having calculated the appropriate weights, one can derive dMOM return in month t with the following formula.

$$R_{dMOM,t} = R_{MOM,t} \cdot w_{dMOM,t} \quad (6)$$

### 3.3 Methods

In this section I will briefly discuss the methods used to analyse the overall performance of the portfolios, to assess if the returns are significant, and to make pairwise comparisons of the portfolios.

To assess the overall performance, standard descriptive statistics will be compared between the portfolios, this includes the mean monthly return, and the monthly volatility or standard deviation. Additionally, the annualized Sharpe ratios will be used to evaluate the risk-adjusted returns of each portfolio. Generally, any Sharpe ratio value above 1 is considered to be acceptable, while any value above 2 would be excellent. Furthermore, the Sharpe ratios are tested using the Gibbons, Ross, and Shanken (GRS) test (1989), to verify whether the portfolios are jointly equal to zero, considering a specified factor model: Fama French 3 Factor (FF3) model (Fama and French, 1993). This is done by testing the joint hypothesis $H_0: \alpha_i = 0 \;\; \forall \;\; dMOM_i$, in other words, it checks whether the portfolios are mean-variance efficient with respect to the FF3 factors. To set up the test, the returns of the strategies are regressed on the FF3 factors based on the following regression equation:

$$r_{FF3} = \alpha + \beta_{RMRF}(r_m - r_f) + \beta_{SMB}(SMB) + \beta_{HML}(HML) + \epsilon \quad (7)$$

where $\alpha$ is the constant, $r_m - r_f$ is the market risk premium, SMB is the historic excess returns of small-cap companies over large-cap companies, HML is the historic excess returns of value stocks (high book-to-price ratio) over growth stocks (low book-to-price ratio), and $\beta_{RMRF}$, $\beta_{SMB}$, and $\beta_{HML}$ are the respective coefficients. The values of the factors used in this calculation are taken from Kenneth French's library. From this regression the betas ($\beta$) and residuals ($\epsilon$) are obtained, for each strategy. Then, the mean excess returns of each portfolio, and the covariance matrix of the residuals are calculated. Finally, the test statistic (GRS) is computed following:

$$\text{GRS} = \frac{(T-N-K)/N}{1+\overline{F^T}\Sigma_F^{-1}\overline{F}}\overline{R^T}\Sigma_\epsilon^{-1}\overline{R} \quad (8)$$

Where $T$ is the number of time periods, $N$ is the number of portfolios, $K$ is the number of factors, $\bar{F}$ is the mean vector of factor returns, $\Sigma_F$ is the covariance matrix of factor returns, $\bar{R}$ is the mean vector of portfolio excess returns, and $\Sigma_\epsilon$ is the covariance matrix of the residuals. The appropriate p-value is derived from an F-distribution N and $T - N - K$ degrees of freedom:

$$p - value_{GRS} = 1 - F(GRS, N, T - N - K) \tag{9}$$

Moreover, skewedness, kurtosis, and the maximum drawdown are considered. When considering skewedness, one would ideally aim for their returns to be positively skewed, indicating that an investor should expect frequent small losses, and few large returns. Regarding kurtosis, high (low) values of kurtosis indicate a higher (lower) chance of extreme outcomes occurring. Most investors regard a combination of positive skewness and moderate kurtosis as optimal. Additionally, the maximum drawdown is an intuitive measure of risk, as it represents the largest month-to-month loss that the portfolio experienced in-sample. Finally, for each portfolio I will calculate the round-trip costs that would render the profits of the strategy insignificant at a significance level of $\alpha = 0.05$, and $\alpha = 0.01$. This is done following Grundy and Martin (2001):

$$Round - Trip\ Cost_{\alpha=5\%} = \left(1 - \frac{1.96}{t - stat_S}\right)\frac{\bar{\mu}_s}{\overline{TO}_s} \tag{10}$$

In the case in which $\alpha = 1\%$, 2.576 is used instead of 1.96. In the above formula $t - stat_S$ is the t-test value of the portfolio returns, and $\bar{\mu}_s$ in the portfolio mean return. On the other hand, based on Barroso and Santa-Clara (2014), $\overline{TO}_s$ is calculated as follows:

$$\overline{TO}_s = 0.5 \cdot \sum_{i}^{N_t} \left| \frac{w_{i,t}}{L_t} - \frac{w_{i,t-1}}{L_{t-1}} \right| \tag{11}$$

In Formula 11 $N_t$ represents the number of stocks at time t, $w_{i,t}$ is the weight inside MOM of stock i at time t, and $L_t$ is the scaling weight of dMOM calculated using Formula 3. It must be mentioned that $w_{i,t-1}$ does account for the return obtained over the holding period as it represents the weight $w_{i,t-1}$ of the previous period multiplied by the returns obtained in the current period $\mu_{1,t} + 1$.

When assessing if the portfolio returns are significant, two measures will be used: a paired t-test with $H_0$: $\bar{\mu} = 0$, and the portfolio returns will be fitted to a (FF3) model (Fama and French, 1993). The t-test is used to observe whether the returns are significantly different from 0. When the p-value of the test is smaller than the chosen confidence level, one can reject the null hypothesis $H_0$, and thus conclude that the mean is significantly different from 0. Differently, fitting the portfolio returns to a FF3 model, is done to verify if the portfolio has a significant constant ($\alpha$). If $\alpha$ is significant, this implies that the model is not able to fully explain the returns of the portfolio, based on the factors of the regression, in this case a FF3 model, based on Formula 7.

Finally, to make pairwise comparisons between the portfolios, I will use both parametric and non-parametric tests. Parametric tests have higher power and are thus more likely to identify a significant effect when one truly exists. However, parametric tests require strong assumptions such as normality of data and might lead to biased results in case these assumptions are not satisfied. Thus, both tests will be used to ensure robustness of results. Additionally, as a high number of t-tests will be used, the risk for false positives is elevated. By using a different pairwise test, one can accept or reject the results of the original t-test, and significantly reduce such risk. The parametric test used in this paper will be a standard paired t-test to compare the mean returns of the two portfolios. The paired t-test has been chosen, as the equality of standard deviations was tested with a variance ratio test, and the results never allowed to reject the null (see Appendix A). The non-parametric test used in this study will be a sign test. The sign test is used to determine if there is a significant difference between the medians of paired or matched samples by comparing the number of positive and negative differences. The test counts the number of times one condition exceeds the other and uses the binomial distribution to evaluate whether the observed difference is statistically significant.

# CHAPTER 4 Results

In this section the performances of the enhanced momentum strategies are compared to each other and to the benchmark strategy MOM. In the first part I will evaluate the companies based on individual metrics, while in the second part the companies' performances will be directly compared to each other. Table 2 includes all the statistics regarding the individual performance of each strategy. Table 3 includes the test statistics of all the pairwise tests performed to evaluate whether one method significantly outperforms the other when used to forecast volatility for a dMOM strategy.

**Table 2: Portfolio Performance Measures.**

| | MOM | BSC | MC | GJR | ML |
|---|---|---|---|---|---|
| **Return Statistics** | | | | | |
| Mean Ret. (in %) | 0.850 | 1.499 | 1.461 | 1.376 | 1.255 |
| | (7.063)* | (12.454)* | (12.133)* | (12.454)* | (12.453)* |
| $\alpha_{FF3}$ (in %) | 0.97* | 1.49* | 1.45* | 1.35* | 1.24* |
| Sharpe Ratio (annualized) | 0.619 | 1.309 | 1.268 | 1.177 | 1.049 |
| **Portfolio Risk Measures** | | | | | |
| Volatility (in %) | 3.255 | 3.257 | 3.257 | 3.257 | 3.257 |
| Skewness | -0.687 | 2.264 | 2.620 | 4.251 | 3.815 |
| Kurtosis | 5.157 | 10.896 | 12.532 | 33.735 | 27.060 |
| Max. Drawdown (in %) | 21.694 | 18.331 | 16.858 | 13.373 | 16.850 |
| **Turnover and Maximum Transaction Costs** | | | | | |
| Mean Turnover (in %) | 53.145 | 63.957 | 61.297 | 63.957 | 63.957 |
| Max. Round-Trip costs | 1.155 | 1.975 | 1.998 | 1.975 | 1.975 |
| | (1.016) | (1.859) | (1.877) | (1.859) | (1.859) |

*Note*: The table includes results of different performance measures: (1) the mean monthly return (in %), (2) the corresponding t-statistic, (3) the constant of the FF3 model fitted to the strategy's returns (in %), (4) the annualized Sharpe ratio, (5) the monthly volatility of returns (in %), (6) the skewness of the return distribution, (7) the kurtosis of the return distribution, (8) the max drawdown of the strategy, (9) the mean turnover in the portfolio between each period, (9) the round-trip costs which render the profits insignificant at $\alpha = 5\%$, and (10) the round-trip costs which render the profits insignificant at $\alpha = 1\%$. MOM is the basic momentum strategy, BSC is the dMOM strategy with the volatility estimated following Barroso and Santa Clara (2014), MC is the dMOM strategy with the volatility estimated using the Monte Carlo approach, GJR is the dMOM strategy with the volatility estimated using a GJR-GARCH model, and ML is the dMOM strategy with the volatility estimated using machine learning.
* Significant at the 1 percent level.

As can be seen in Table 2, the standard momentum factor (MOM) returned an average of 0.85% monthly between January 1962 and June 2024. This value is statistically significant at a confidence level of $\alpha = 0.01$ having a t-statistic of 7.063. Additionally, the strategy has a statistically significant FF3 $\alpha_{FF3} = 0.010$, implying that the strategy generates, monthly, a 0.97% return which cannot be explained by the FF3 factors (RMRF, SMB, and HML). However, as indicated by other authors (Barroso and Santa-Clara, 2015; Daniel and Moskowitz, 2016), the MOM strategy has high kurtosis and negative skewness which leads to high maximum drawdowns. These results replicate in my study and are visualized in Figure 1. Moreover, partially due to the fat left tails, the strategy has a sub-optimal Sharpe ratio of 0.619. However, due to the conservative approach used to filter the data, the sample likely has a reduced volatility compared to the samples of the other authors, thus these results are not particularly extreme. This can be verified by the fact that the MOM strategy has a 21.694% maximum monthly drawdown in the period, which is significantly lower than what other authors report (Hanauer and Windmuller, 2022). Regardless, this is a significantly high value, as it must be kept in mind that it is expressed in monthly terms and would correspond to a 75% drawdown when the value is annualized. Finally, it must be mentioned that the strategy has a considerable amount of turnover, as on average, more than 53% of the positions change monthly, which, combined with the high returns, results in a maximum acceptable transaction cost just over 1%.

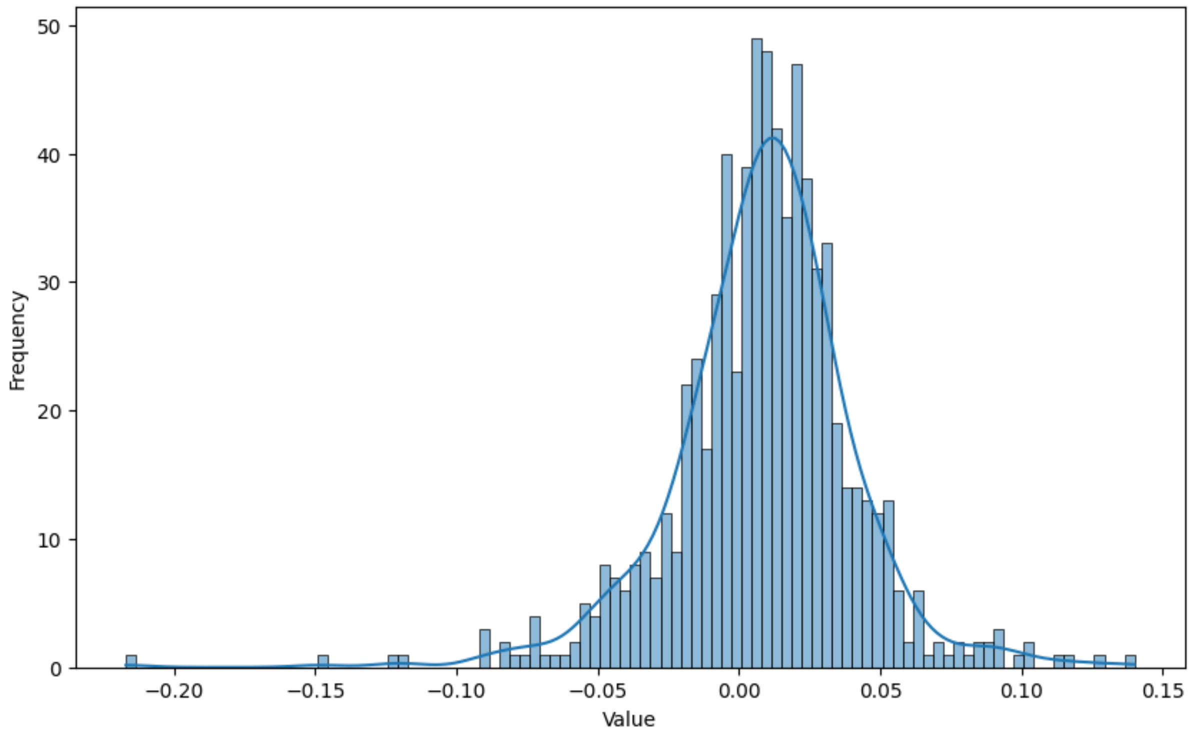


**Figure 1: Distribution of MOM returns.**

This figure displays a histogram, portraying the frequency of each value of returns of MOM over the whole sample, with a distribution curve. The values of "Values" are in USD.

When considering the enhanced momentum strategies, a significantly better performance is observed compared to MOM. All dMOM strategies outperform MOM in every metric, except for their increased kurtosis. Overall, the volatility, the mean turnover, and the maximum round-trip costs appear to be more or less equivalent for the 4 dMOM strategies using the Barroso and Santa-Clara (2015) method (BSC), the Monte Carlo approach (MC), a GJR-GARCH model (GJR), and the Machine Learning approach (ML). Similarly, when considering mean returns, the 4 dMOM strategies have extremely significant t-statistics, signifying that the returns are very unlikely to be 0. Alongside being significant, the mean returns are also positive and considerably larger in scale compared to those of MOM, and range between 1.49 % (BSC) and 1.24% (ML) per month. Additionally, all dMOM strategies have statistically and economically significant values of $\alpha_{FF3}$ which implies that the strategies generate unexplained returns. All dMOM strategies have a Sharpe ratio above 1 which implies that the portfolios yield more than acceptable risk-adjusted return. In this metric, the BSC strategy clearly has the best performance, under all metrics, with a standout 1.309 Sharpe ratio indicating the strategy has the best risk weighted returns in the sample. Moreover, the GRS test statistic is 20.765, which corresponds to a p-value of $1.11 \cdot 10^{-16}$ (see Appendix B). The extremely low p-value suggests that at least one of the strategies has a Sharpe ratio that is significantly different from the others when adjusted for the Fama-French three factors, leading to the rejection of the null hypothesis of equal Sharpe ratios. This implies that there are significant differences in the Sharpe ratios of the five strategies tested, and that the overperformance of BSC, as goes for risk-adjusted returns, is indeed significant. An interesting thing to notice is that MC trails BSC in all categories, and has similar statistics, which is an argument in favor of Hypothesis 1, which states that MC is not an improved version of BSC. As for returns, the GJR and the ML strategies resulted in the worst overall performances. This is especially clear when looking at Figure 2 which visually compares the cumulative returns of the 4 dMOM strategies, MOM, and the market. The 5 strategies outperform the market; however, MOM's outperformance is not as marked.

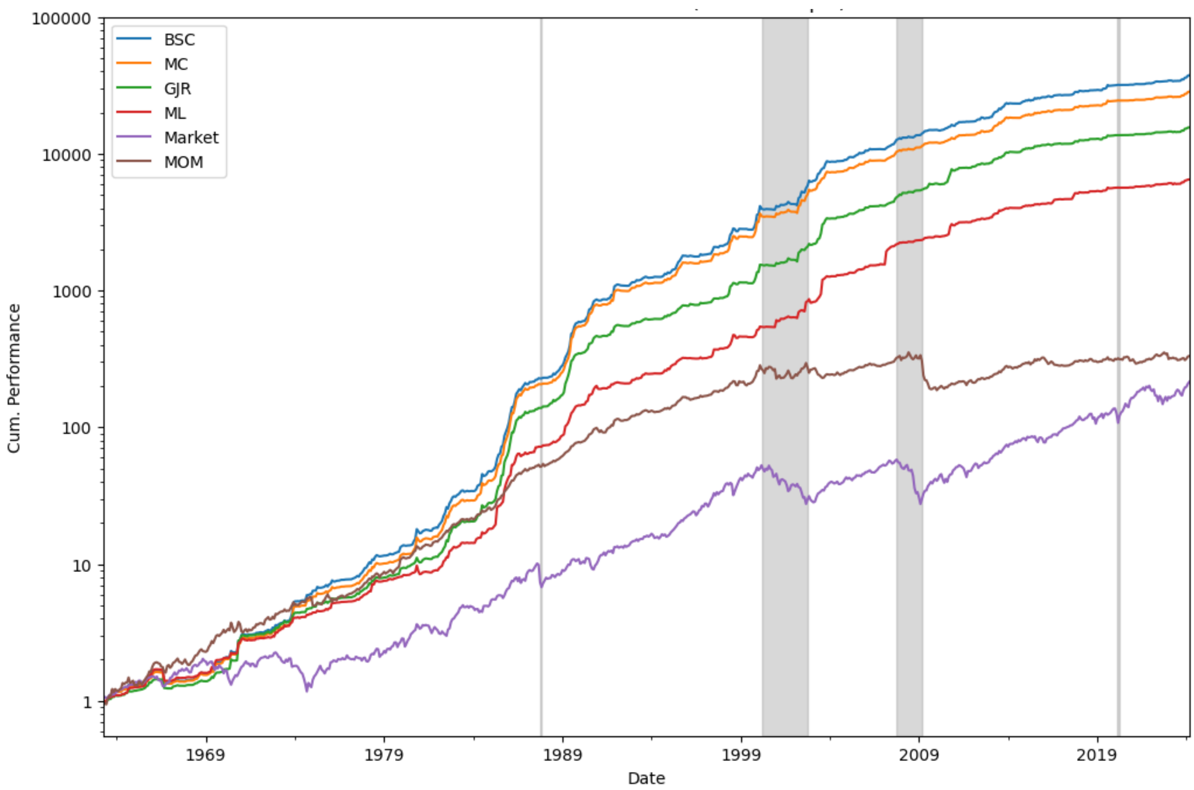


**Figure 2: Cumulative performance of the momentum strategies and the market.**

The figure displays the cumulated performance of a \$1 investment in each of the momentum strategies (the risk-free rate is included as they are long-short strategies) and the market. The following strategies are included: MOM, BSC, MC, GJR, and ML. MOM is the standard momentum strategy, the other are dMOM strategies with volatilities calculated using respectively the mean variance (BSC), a Monte Carlo simulation (MC), a GJR-GARCH (GJR), and machine learning (ML). The shaded areas of the graph represent major financial crises, from left to right: Black Monday (1987), Dot-Com Bubble Burst (2000-2002), 2008 Financial Crisis, and COVID-19 Market Crash. For details regarding variable construction, see Section 3.2. The sample period ranges from 01/1962 to 06/2024.

From a portfolio risk management perspective, by looking at Table 2 it is not clear which strategy has the best performance. As previously mentioned, all dMOM strategies have the same value for the standard deviation, when rounding to the third decimal. However, the standard deviation is not the best indicator when evaluating portfolio risk, and this does not allow us to conclude that the strategies are equivalent. When considering skewness, the dMOM strategies significantly outperform the standard MOM strategy, as they have high, positive skewness compared to the negative skewness of MOM. As the dMOM strategies have positive skewness and mean $\bar{\mu}_s > 0$, extreme losses are very unlikely, values close to or slightly under the mean are the most common, and few extremely high gains are to be expected. This makes the strategies' performance, particularly attractive. GJR exhibits the highest skewness at 4.251, with ML following closely at 3.815. As for the previous discussion related to return metrics, BSC and MC have similar values for skewness.

When considering kurtosis, my results do not replicate Hanauer and Windmuller (2022), as the dMOM strategies have considerably higher kurtosis than MOM and their returns' distributions are leptokurtic. The authors find that, in their sample, kurtosis is reduced by volatility scaling. Moreover, these values are significantly high and range between 33.735 (GJR) and 10.896 (BSC). High levels of kurtosis are not a particular concern in this case as it implies that the distributions have fat tails. In other words, extreme values (both on the left and right side) are considerably more common. However, as the distributions of dMOM returns are positively skewed and have mean $\bar{\mu}_s > 0$, high kurtosis could be a positive characteristic as extreme positive values are more likely. Additionally, it can be verified that this is not an issue by observing the maximum drawdown of the strategies in Table 2, as the enhanced momentum strategies have lower values for this statistic compared to MOM. BSC has the highest maximum drawdown at 18.33%, which is 3% less than MOM, whereas GJR has the lowest maximum drawdown at 13.37%, 8% lower than MOM. These values provide tangible evidence that volatility scaling reduces the overall risk exposure of the MOM. The data structures of the distributions of dMOM strategies can be better observed in Appendix C.

Moreover, by looking at Figure 2, one can observe that, differently from MOM and the market, the dMOM strategies are robust to the major crises (shaded areas). Studying the performance of these strategies during crises is particularly important, as such events constitute the greatest downwards risk for most portfolios. Most strategies, incur sizeable losses during market crashes, as can be seen from the performance of MOM. However, the dMOM strategies appear significantly resilient and even be able to capitalize on the reversal, unlike MOM which is particularly sensitive to such events. This is due to the fact that the portfolio weight of dMOM partially depends on MOM's forecasted return, which can be both positive and negative. Thus, dMOM weights are negative in periods of crisis causing the strategy to be inverted (long losers, and short winners).

Finally, when comparing the strategies based on the maximum acceptable transaction costs, and mean turnover, it can be observed that the dMOM strategies have similar values in these metrics (Table 2). The dMOM strategies have an overall higher average turnover, compared to MOM; this difference is amounts to roughly 10%. However, the maximum transaction costs to render profits insignificant are surprisingly higher at both levels of $\alpha$ ($\alpha = 0.05$ and $\alpha = 0.01$), when compared to MOM. This indicates that the strategies are profitable in some cases in which MOM fails to generate profit. This is likely driven by the significantly higher returns of the dMOM strategies, which overcompensate for the higher overall amount of turnover.

**Table 3: Pairwise tests' p-values.**

| | MOM | BSC | MC | GJR | ML |
|---|---|---|---|---|---|
| MOM | | 0.000 | 0.000 | 0.000 | 0.004 |
| BSC | 0.042 | | 0.000 | 0.036 | 0.007 |
| MC | 0.112 | 0.000 | | 0.155 | 0.024 |
| GJR | 0.853 | 0.000 | 0.024 | | 0.146 |
| ML | 0.130 | 0.001 | 0.007 | 0.171 | |

*Note*: In the table above one can find the p-values of paired t-tests (parametric; in the dark grey section), and of sign tests (non-parametric; in the white section). All tests are two-sided. Tests comparing the returns of the following strategies are included: MOM, BSC, MC, GJR, and ML. MOM is the standard momentum strategy, the other are dMOM strategies with volatilities calculated using respectively the mean variance (BSC), a Monte Carlo simulation (MC), a GJR-GARCH (GJR), and machine learning (ML). For details regarding variable construction, see Section 3.2.

In Table 3, one can observe the results of the pairwise tests performed between the 5 strategies (MOM + 4 dMOM). Interestingly, the results of the t-test and sign test, contrary to expectations, indicate that returns of BSC and MC are significantly different at $\alpha = 0.01$. While these results appear to be in contrast with the conclusions derived from Table 2, using statistical tests to compare two strategies is a more robust method of analysis. This is due to the fact that statistical tests can identify small differences which would be overlooked when comparing performance metrics just by magnitude. On the other hand, both the t-test and the sign test do not allow to reject the null hypothesis that the returns of GJR and ML are statistically different, even at a confidence level of $\alpha = 0.1$. This result is striking as the difference in performance appeared to be more marked by just looking at Table 2. Thus, this finding suggests that the machine learning model, at the base of ML, strongly bases its forecasts on the forecasts of GJR-GARCH models which are present in its training dataset. Differently, the t-test results allow to reject the hypothesis that the mean returns are different only at $\alpha = 0.05$ when comparing BSC to GJR, and MC to ML. In both cases, the sign test allows rejection at $\alpha = 0.01$. In cases in which the two statistical tests lead to opposite conclusions the interpretation is not straight forward. The t-test, being parametrical, has higher power, which implies that it is more likely to identify an effect in the case that an effect exists, when compared to the sign test. However, the t-test requires strong assumptions on the distribution of data, more specifically, that data is normally distributed. Based on the results in Table 2, and by looking at Appendix C, one can observe that the distributions of returns of the dMOM strategies are not normal, thus the normality assumption is not satisfied. This condition is likely to lead to biased t-test results. Thus, one can conclude that in this case, the sign test is more reliable, and that BSC and GJR, and MC and ML, most likely have statistically different means at $\alpha = 0.01$. Similarly, in the case of the pairwise tests between MC and GJR, the t-test does not allow to reject the null at any significance level $\alpha \leq 0.1$, however the sign test rejects the null at $\alpha = 0.05$. In this case the result of the sign test is deemed to be more appropriate. In conclusion, the returns of all strategies are statistically different, at a confidence level of at least $\alpha = 0.05$, except for the returns of GJR and ML which are not statistically different from each other.

Many authors have found that markets behave differently in discrete periods of time such as Perron (1989) who argues that major economic shocks have lasting effects on financial markets. Additionally, one must consider that cumulative returns are intrinsically strongly dependent on the performance at the start of the sample. Due to the variability of stock markets, and the nature of cumulative returns, a graph is included tracking the cumulative performance of the strategies in the past 20 years (Figure 4). Figure 4 paints a completely different picture of the strategies' performance. To start, it must be noted that the market outperforms all strategies in this sample period. Additionally, it should be noted that MOM exhibits poor overall performance, as it never fully recovers from the 2008 financial crisis. This is reflected in the performance of the dMOM strategies which fail to outperform the market, while considerably outperforming the standard MOM. Finally, one can notice that ML surprisingly has the best performance, followed GJR, BSC, and MC. These results are in contrast with those obtained when considering the full sample, in which BSC and MC outperformed the other strategies.

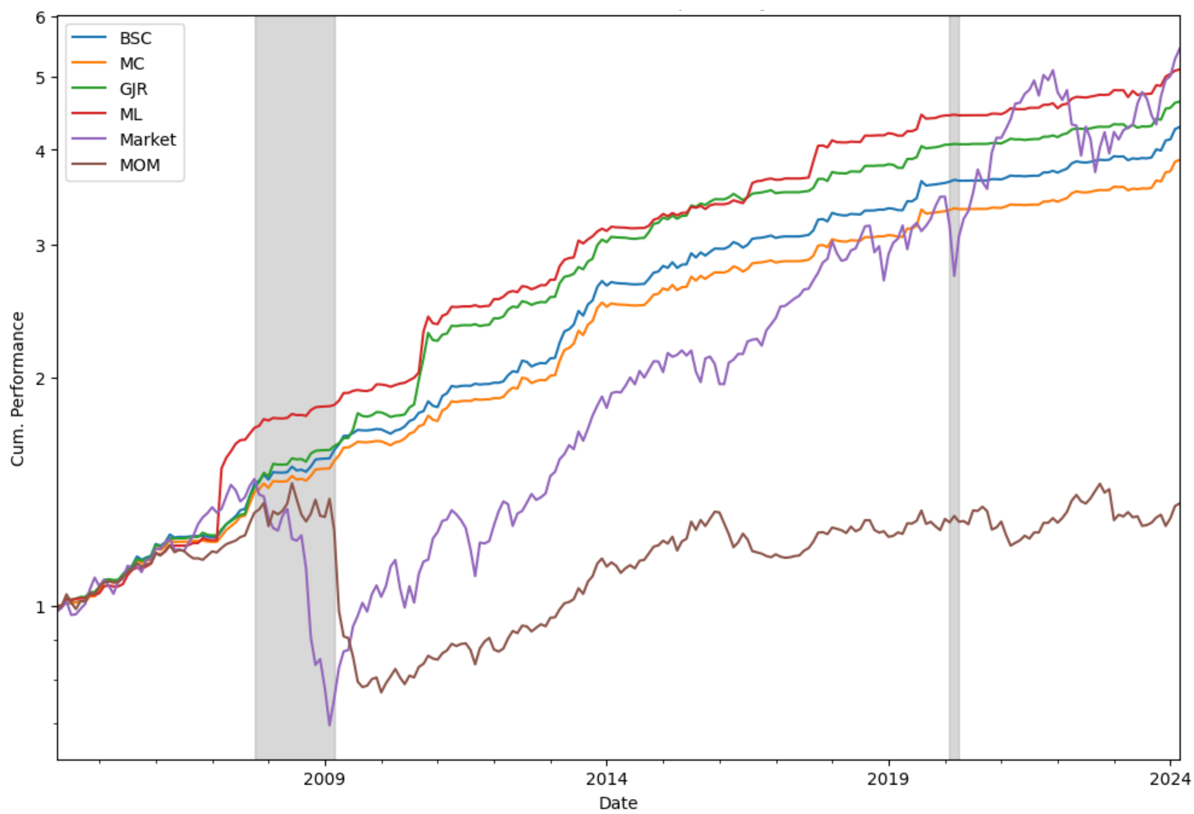


**Figure 4: Cumulative performance of the momentum strategies and the market 2004-2024.**

The figure displays the cumulated performance of a $1 investment in each of the momentum strategies (the risk-free rate is included as they are long-short strategies) and the market. The following strategies are included: MOM, BSC, MC, GJR, and ML. MOM is the standard momentum strategy, the other are dMOM strategies with volatilities calculated using respectively the mean variance (BSC), a Monte Carlo simulation (MC), a GJR-GARCH (GJR), and machine learning (ML). The shaded areas of the graph represent major financial crises, from left to right: 2008 Financial Crisis, and COVID-19 Market Crash. For details regarding variable construction, see Section 3.2. The sample period ranges from 06/2004 to 06/2024.

# CHAPTER 5 Discussion

This section will discuss the results presented in the "Results" section and compare them to previous literature when possible. Additionally, the results will be discussed in the optic of the 3-hypothesis stated at the start of the paper.

To start, the most important finding at the base of this study is that volatility scaling, in the form of dynamic-scaled momentum, does in fact improve the performance of the standard MOM. These results are almost perfectly in line with those of Daniel and Moskowitz (2016), and Hanauer and Windmuller (2022). Both papers find that the dMOM strategies have significant and increased mean return (compared to MOM), significant and unexplained excess returns ($\alpha_{FF3} > 0$), lower maximum drawdown, and positive skewness. However, an aspect that did not replicate, in this study, is the lower level of kurtosis. In this analysis, dMOM strategies have considerably higher levels of kurtosis compared to MOM. This is likely due to the differences in the samples and is not a major concern, as all other aspects replicate. The authors use data starting in 1930 and ending in 2017, while my study uses data starting in 1962 and ending in 2024. Additionally, my sample likely has a significantly lower volatility compared to that of Hanauer and Windmuller (2022), who report an annualized in-sample volatility of MOM of $\sigma_{MOM} = 19\%$, while in my case it is $\sigma_{MOM} = 11\%$. As previously stated, this difference is most likely driven by the conservative approach used in data cleaning processes in this study. Additionally, the authors use a sample of over 23,000 securities which leads their MOM portfolio to be larger in size and have lower maximum transaction costs at the same confidence levels due to the higher absolute turnover.

When comparing the performance of the dMOM strategies (BSC, MC, GJR, and ML) to each other, the interpretation is not particularly clear. From the results of pairwise testing in Table 3, one can observe that all strategies have returns which significantly different from each other, except for GJR and ML. However, as discussed in Section 4, this is due to the fact that the training dataset of the machine learning model at the base of ML has all the forecasts used to estimate GJR, and the other models. This implies that the volatility forecast used in ML is based on the forecasts used in the other strategies, and especially on those of GJR (based on the ML portfolio performance). The fact that the returns of the strategies are significantly different implies that BSC yields the highest return on average, followed by MC, and GJR and ML, over the full sample. However, when the most recent (20 year) sample is considered, ML yields the best returns, followed by GJR, BSC, and MC. By observing these rankings (for returns in both the full and shortened sample), one can conclude that MC is a strictly dominated strategy by BSC, as goes for returns. In the full sample, BSC and MC are the best strategies based on return metrics. Nonetheless, when assessing the performance of these strategies from a risk management perspective, they have the highest maximum drawdowns in the sample, and the lowest values for skewness. This implies that BSC and MC are overall riskier than GJR and ML. On the other hand, it must be considered, that, based on the Sharpe ratios (which are proven to be statistically different), BSC and MC offer better reward for the risk taken, which is likely the reason why they outperform in the full sample.

The first hypothesis (**H1**), states that the Monte Carlo (MC) approach does not improve the portfolio's performance when compared to the mean volatility model (BSC). The expectation is based on the fact that due to the law of large numbers, if a sufficiently large number of simulations is performed, the historical mean variance should be equal to the variance forecasted with the Monte Carlo approach. This is expected even if the Monte Carlo approach has the advantage of being a forward-looking approach. Thus, according to expectations BSC and MC should have had similar performances in all metrics, and the 2 strategies should have had the same mean return according to pairwise testing. As goes for performance, BSC outperforms MC for all return metrics in all samples, and MC outperforms BSC from a risk-management perspective, having a lower maximum drawdown (by 1.5%), and a higher positive value for skewness (2.620 compared to 2.264). However, these differences are marginal, and one could argue that this condition was satisfied, especially when considering the cumulative performance. On the other hand, the pairwise testing did not return the expected results, as the mean strategy returns are statistically different at a level of alpha $\alpha = 0.01$. Since the strategies are indeed different, **H1** cannot be accepted or rejected. Overall, the strategies are statistically different, and there is no universally better option. More risk averse investors are likely to find MC more attractive than BSC due to its better risk management capabilities, while return-optimizing investors would prefer BSC.

Hypothesis 2 (**H2**) states that the GJR-GARCH volatility estimation technique (GJR) improves the portfolio's performance when compared to the mean volatility model (BSC). This hypothesis was based on the fact that the volatility forecasting approach of BSC (mean historical volatility), does not account for the autocorrelation present in time-series data. This implies that the GJR-GARCH volatility forecasts are supposed to be more accurate. However, it must be noted that the most accurate volatility forecasts do not necessarily lead to the best portfolio performance, as one cannot conclude that the dynamic volatility scaling process benefits from the added precision. In this case, when considering the full sample, GJR had worse mean returns, lower $\alpha_{FF3}$, and lower Sharpe ratio than BSC. Yet, in the most recent (20 year) sample, GJR has the second-best cumulative performance, outpacing BSC. This difference in performance is most likely due to GJR's better risk management. The past 20 years saw two of the largest financial crises in history: the 2008 financial crisis, and the COVID-19 crisis (2019), which cause a flat performance of BSC and generated speculation opportunities for GJR, which spikes during market reversals. Additionally, GJR completely outperforms all strategies (BSC in particular) as goes for risk management. This is particularly clear when considering that GJR has almost twice the value for (positive) skewness as BSC, and a maximum drawdown which is reduced by 5%. In this case, differently from **H1**, the improvement in risk-management cannot be ignored, however also the difference in mean returns over the whole sample is sizeable and BSC has a higher Sharpe ratio. For these reasons, similarly to **H1**, also **H2** cannot be accepted or rejected as more risk seeking investors would find BSC more appealing regardless of the significantly better performance of GJR in risk management and during economically unstable periods. This conclusion is also in contrast with the expectations of Hanauer and Windmuller (2022), who state that by controlling for the positive autocorrelation, one can generate scaling weights that increase the Sharpe ratio of enhanced momentum strategies.

Finally, the third and last hypothesis (**H3**), states that the machine learning algorithm yields the greatest improvement to the portfolio, compared to the other methods. The expectation is based on the fact that the machine learning model has all forecasts by the other models in its training dataset, plus moving averages and other data. This implies that the model makes forecasts based on forecasts of other models and additional data and is combining the intuitions of the other models. Due to this, the volatility forecasts used to calculate ML were expected to be the most precise. When considering ML's performance, it is trailing all other enhanced portfolios under every return metric and has the worst cumulative performance in the full sample out of the dMOM strategies. However, when considering the risk-related metrics, ML, has the second-best performance very close to that of GJR. The overall similarity to GJR is confirmed by both pairwise tests which do not allow to reject the hypothesis that the mean returns of the two models are identical. This implies that the predictions of ML are strongly based on the volatility forecasts made with the GJR-GARCH model.

On the other hand, ML is the best performer in the most recent (20 year) sample. This alone is not enough to state that ML had the best performance out of all the dMOM strategies, considering that it lags GJR in every aspect in the full sample. As mentioned during the discussion of **H2**, it must be kept in mind that the most accurate forecast is not necessarily the optimal one for the dynamically scaled momentum strategy. Thus, ML's underperformance is not necessarily indicative of imprecise forecasts. Additionally, another source of bias is using squared returns (variance proxy), as a target value for the model. This implies that the machine learning model will use its data to make a forecast, as close as possible to the realized squared returns, which is only a proxy for the actual realized variance. Thus, the model is trained to target an approximation of the real value, which inevitably introduces a strong bias in the estimation process. All considered, it would be incorrect to state that ML is the best model as it is outperformed by BSC from a returns and Sharpe ratio perspective, and by GJR from a risk management perspective. This implies that no rational investor would choose ML over BSC or GJR. Thus, the third hypothesis (**H3**) is rejected.

In conclusion, BSC is the best performing model for risk-seeking investors, which aim to maximize their returns and/or their portfolio's Sharpe ratio. On the other hand, GJR is the preferred strategy for risk-averse investors, due to its superior risk management. Finally, one can conclude using more sophisticated models to forecast volatility helps significantly reduce portfolio risk, which is the primary objective of volatility scaling.

# CHAPTER 6 Conclusion

This study explores the impact of various volatility forecasting methods on the performance of dynamically scaled momentum strategies within the U.S. equities market. Historically, momentum strategies have consistently outperformed the market, and yielded substantial returns, however they are particularly sensitive to market reversals, during which they experience significant drawdowns. Such drawdowns make the momentum strategy especially unattractive to risk-averse investors, or investors with tight budget constraints (due to the risk of margin calls). To address this issue, multiple approaches have been put forward by researchers, a particularly interesting one being volatility scaling. In this paper, a specific type of volatility scaling is studied: dynamically scaled momentum. Volatility scaling, in general, aims to improve a momentum portfolio's performance by accounting for the leverage effect. More specifically, the dynamically scaled momentum strategy is a combination of volatility scaling and a dynamic Markowitz model, which allows the weights of the portfolio to vary (be positive or negative) based on market conditions and return forecasts. In this study, the performance of different volatility forecasting techniques is analysed alongside their effects on the performance of the strategy. Using the historical mean variance (BSC) for volatility scaling, has been empirically proven to work, however this study tested whether more advanced techniques, such as Monte Carlo simulations, GJR-GARCH models, and machine learning algorithms could further improve the strategy's performance. This paper provides a comparison of these methods. Thus, the study's main purpose was to understand how risk can be better managed when investing using a momentum strategy. Consequently, the main research question of this paper was: "How do different volatility forecasting methods impact the performance of dynamically scaled momentum strategies in the U.S. equities market?"

To compare the portfolio performances, in order to be able to address the research question, the analysis starts with the construction of a long-short momentum portfolio. To apply volatility scaling, the appropriate scaling weights were estimated using different forecasting methods, to be then applied to MOM. Firstly, the mean historical return (BSC) was used, a simple but effective approach used in previous research. This strategy was used as the benchmark, to test whether more advanced volatility forecasting techniques could bring an improvement to the strategy. Then, Monte Carlo simulations were used to generate forward-looking volatility data for the forecasts of at the base of MC. Additionally, the GJR-GARCH model was used to make volatility forecasts for GJR, as the method accounts for the positive autocorrelation of data, and for the inherent asymmetries of stock market volatility. Finally, machine learning algorithms were used make volatility forecasts to estimate weights for ML, such models have the advantage of not requiring initial assumptions. The performance of these portfolios was then evaluated and compared by considering both return-based and risk-based metrics. This double approach allows to have a better overview of the strategies' performances. Moreover, pairwise tests were performed to verify whether the portfolios had statistically different mean returns.

Finally, the results indicate that the dynamic momentum strategies: BSC, MC, GJR, and ML consistently outperform the traditional MOM strategy across various metrics in the sample. The BSC method yielded the highest mean returns and Sharpe ratio, making it particularly attractive for risk averse or risk-weighted-return maximizing investors. While the GJR-GARCH model had the best performance with regards to risk management, due to it having the lowest maximum drawdowns and highest positive skewness. These characteristics make the strategy particularly attractive to risk-averse investors. The analysis of the past 20 years reveals a shift in the performance dynamics of these strategies. Purely based on cumulative performance, ML outperforms its peers when using the past 20 years (June 2004, to June 2024) as a sample. This was due to the model's high resilience during financial crises, and its ability to rebound during market reversals which lead to spikes in its performance. This proves its value and efficiency during periods of market uncertainty and periods prone to crises. However, no strategy was found to consistently outperform the other. Overall, this research highlights the importance of adapting investment strategies to different market conditions and risk appetites and provides further empirical evidence of the dMOM strategies' outperformance compared to the standard MOM, and the market.

Concluding, the more advanced and complete volatility estimation methods significantly improve the performance of momentum strategies from a risk management perspective, by reducing overall risk exposure. However, simpler volatility forecasting techniques, such as using the historical mean variance, appear to be almost as effective at hedging for risk, while outperforming from a returns and risk-weighted returns perspective. Thus, no universally better volatility forecasting method was identified. As financial markets continue to evolve, the application of more advanced models and techniques will be vital in developing new strategies to deal with the modern problems of financial markets.

The main limitations of this study include the underlying subjectivity of the performance evaluation process, and the limits to practical applicability of these methods. The former limitation of this study is clear, when considering that 2 out of 3 hypotheses could neither be rejected nor accepted. This would not be a problem in cases like Hypothesis 3 (**H3**) in which the performance of one portfolio is strictly dominated. Unfortunately, in most cases the interpretation is ambiguous. This is due to the risk-preferences of investors and no universally accepted framework exists to correct for this. A potential solution would be to choose one type of investor (risk-seeking or risk-averse) and make the portfolio evaluation based on their preferences. However, this approach does not consider the portfolio's performance as a whole and might lead to sub-optimal conclusions. When considering the limits to the practical applicability of these methods, one must consider the complexity of calculations, the large portfolio size (which has a maximum size of 1,380 stocks), the fact that an investor must change the portfolio almost fully (average turnover >60%) every month, transaction costs, and the constraints related to short selling. These factors (except the limits to short selling) could be mitigated, by reducing the sample of stocks, however this exposes the investor to higher variability depending on the sample size and constituents. Unfortunately, the limitations related to short selling cannot be ignored, as this strategy could not be modified to be long-only. Regardless, dMOM strategies are at least as implementable as MOM, and tend to be profitable even with higher levels of transaction costs.

Further research should consider more efficient volatility forecasting methods such as more sophisticated machine learning techniques, to better identify patterns in the data, and improve the quality of predictions. Additionally, setting an optimized value of portfolio performance as a target for the model, instead of the realized volatility, might help output volatility forecasts that better suit the model. Moreover, better volatility proxies can be used, in order to further improve the performance of dMOM strategies and enhance their risk management capabilities. On the other hand, improved versions of dMOM could arise, which consider indicators such as market sentiment, or tackle liquidity issues inherent to some securities. Such models could further improve the performance of MOM due to their better understanding of real market conditions. These advancements would help further commercialize momentum strategies and make them an investment vessel that caters to any type of investor.

# REFERENCES


*A Test of the Efficiency of a Given Portfolio on JSTOR*. (n.d.). Retrieved June 25, 2024, from https://www-jstor-org.eur.idm.oclc.org/stable/1913625

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, *96*(453), 42–55. https://doi.org/10.1198/016214501750332965

Asem, E., & Tian, G. Y. (2010). Market Dynamics and Momentum Profits. *Journal of Financial and Quantitative Analysis*, *45*(6), 1549–1562. https://doi.org/10.1017/S0022109010000542

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and Momentum Everywhere. *The Journal of Finance*, *68*(3), 929–985. https://doi.org/10.1111/jofi.12021

Bachelier, L. (1900). Théorie de la spéculation. *Annales scientifiques de l'École normale supérieure*, *17*, 21–86. https://doi.org/10.24033/asens.476

Barroso, P., & Santa-Clara, P. (2014). *Momentum Has Its Moments* (SSRN Scholarly Paper 2041429). https://doi.org/10.2139/ssrn.2041429

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, *81*(3), 637–654. https://doi.org/10.1086/260062

Blitz, D., & van Vliet, P. (2008). *Global Tactical Cross-Asset Allocation: Applying Value and Momentum Across Asset Classes* (SSRN Scholarly Paper 1276616). https://papers.ssrn.com/abstract=1276616

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, *52*(1), 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Chabot, B., Ghysels, E., & Jagannathan, R. (2014). *Momentum Trading, Return Chasing, and Predictable Crashes* (Working Paper 20660). National Bureau of Economic Research. https://doi.org/10.3386/w20660

Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realized volatility1. *Journal of Financial Economics*, *50*(2), 125–150. https://doi.org/10.1016/S0304-405X(98)00034-8

Daniel, G., Sornette, D., & Wohrmann, P. (2008). *Look-Ahead Benchmark Bias in Portfolio Performance Evaluation* (arXiv:0810.1922). arXiv. https://doi.org/10.48550/arXiv.0810.1922

*Definition of MOMENTUM*. (2024, June 2). https://www.merriam-webster.com/dictionary/momentum

Diane, L., & Brijlal, P. (2024). Forecasting Stock Market Realized Volatility using Random Forest and Artificial Neural Network in South Africa. *International Journal of Economics and Financial Issues*, *14*(2), Article 2. https://doi.org/10.32479/ijefi.15431

Dobrynskaya, V. (2019). Avoiding momentum crashes: Dynamic momentum and contrarian trading. *Journal of International Financial Markets, Institutions and Money*, *63*, 101141. https://doi.org/10.1016/j.intfin.2019.101141

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987–1007. https://doi.org/10.2307/1912773

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Glasserman, P. (2003a). Applications in Risk Management. In P. Glasserman (Ed.), *Monte Carlo Methods in Financial Engineering* (pp. 481–537). Springer. https://doi.org/10.1007/978-0-387-21617-1_9

Glasserman, P. (2003b). Applications in Risk Management. In P. Glasserman (Ed.), *Monte Carlo Methods in Financial Engineering* (pp. 481–537). Springer. https://doi.org/10.1007/978-0-387-21617-1_9

Grundy, B. D., & Martin, J. S. M. (2001). Understanding the Nature of the Risks and the Source of the Rewards to Momentum Investing. *The Review of Financial Studies*, *14*(1), 29–78. https://doi.org/10.1093/rfs/14.1.29

Han, Y., Zhou, G., & Zhu, Y. (2016). *Taming Momentum Crashes: A Simple Stop-Loss Strategy* (SSRN Scholarly Paper 2407199). https://doi.org/10.2139/ssrn.2407199

Hanauer, M. X., & Windmüller, S. (2023). Enhanced momentum strategies. *Journal of Banking & Finance*, *148*, 106712. https://doi.org/10.1016/j.jbankfin.2022.106712

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, *48*(1), 65–91. https://doi.org/10.2307/2328882

Jensen, M. C. (1968). The Performance of Mutual Funds in the Period 1945-1964. *The Journal of Finance*, *23*(2), 389–416. https://doi.org/10.2307/2325404

Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate Sustainability: First Evidence on Materiality. *The Accounting Review*, *91*(6), 1697–1724. https://doi.org/10.2308/accr-51383

Kolari, J. W., Liu, W., & Pynnönen, S. (2023). Portfolio Performance Measures. In J. W. Kolari, W. Liu, & S. Pynnönen (Eds.), *Professional Investment Portfolio Management: Boosting Performance with Machine-Made Portfolios and Stock Market Evidence* (pp. 97–119). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-48169-7_6

Larsen, G. A., & Larsen, E. P. (2018). *A Short Note on the Potential for a Momentum Based Investment Strategy in Sector ETFs*. https://hdl.handle.net/1805/21657

Mandelbrot, B. B. (2013). *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. Selecta Volume E*. Springer Science & Business Media.

Marhfor, A. (2016). Portfolio Performance Measurement: Review of Literature and Avenues of Future Research. *American Journal of Industrial and Business Management*, *6*(4), Article 4. https://doi.org/10.4236/ajibm.2016.64039

Markowitz, H. (1952). Portfolio Selection*. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Merton, R. C. (1973). Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, *4*(1), 141–183. https://doi.org/10.2307/3003143

Metropolis, N., & Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, *44*(247), 335–341. https://doi.org/10.1080/01621459.1949.10483310

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, *104*(2), 228–250. https://doi.org/10.1016/j.jfineco.2011.11.003

Nugroho, D. B., Kurniawati, D., Panjaitan, L. P., Kholil, Z., Susanto, B., & Sasongko, L. R. (2019). Empirical performance of GARCH, GARCH-M, GJR-GARCH and log-GARCH models for returns volatility. *Journal of Physics: Conference Series*, *1307*(1), 012003. https://doi.org/10.1088/1742-6596/1307/1/012003

Perron, P. (1989). The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. *Econometrica*, *57*(6), 1361–1401. https://doi.org/10.2307/1913712

Quant Momentum Trade Is Hot Again in Crypto. (2024, April 3). *Bloomberg.Com*. https://www.bloomberg.com/news/articles/2024-04-03/quant-momentum-trade-is-hot-again-in-crypto-amid-sharp-reversal

Roy, A. D. (1952). Safety First and the Holding of Assets. *Econometrica*, *20*(3), 431–449. https://doi.org/10.2307/1907413

Schwert, G. W. (1989). Why Does Stock Market Volatility Change Over Time? *The Journal of Finance*, *44*(5), 1115–1153. https://doi.org/10.1111/j.1540-6261.1989.tb02647.x

Sharpe, W. F. (1966). Mutual Fund Performance. *The Journal of Business*, *39*(1), 119–138.

Wang, C. D., & Mykland, P. A. (2014). The Estimation of Leverage Effect With High-Frequency Data. *Journal of the American Statistical Association*, *109*(505), 197–215. https://doi.org/10.1080/01621459.2013.864189

*What Is Stock Market Volatility? – Forbes Advisor*. (n.d.). Retrieved June 5, 2024, from https://www.forbes.com/advisor/investing/what-is-volatility/

# APPENDIX A: Variance Ratio Test

**Table A1: Variance ratio test.**

| | MOM | BSC | MC | GJR | ML |
|---|---|---|---|---|---|
| MOM | | | | | |
| BSC | 0.985 | | | | |
| MC | 0.985 | 1.000 | | | |
| GJR | 0.985 | 1.000 | 1.000 | | |
| ML | 0.985 | 1.000 | 1.000 | 1.000 | |

*Note*: In the table above one can find the p-values of the variance ratio tests used to compare the distributions of the samples. All tests are two-sided. Tests comparing the returns of the following strategies are included: MOM, BSC, MC, GJR, and ML. MOM is the standard momentum strategy, the other are dMOM strategies with volatilities calculated using respectively the mean variance (BSC), a Monte Carlo simulation (MC), a GJR-GARCH (GJR), and machine learning (ML). For details regarding variable construction, see Section 3.2

## APPENDIX B: GRS Test

**Table B1: Gibbons, Ross, and Shanken (GRS) Test Results and Parameters.**

| | Test Statistic | T | N | K | P-Value |
|---|---|---|---|---|---|
| GRS Test | 20.765 | 732 | 5 | 3 | 1.110e-16 |

*Note*: In the table above one can find the test results and parameters of the GRS test, used to verify whether the portfolio's Sharpe ratios are significantly different. The test statistic is obtained with Formula 8, "T" is the number of observations per portfolio, "N" is the number of portfolios, "K" is the number of factors (of FF3), and "P-Value" is the relevant p-value based on Formula 9. The test compares the following strategies: MOM, BSC, MC, GJR, and ML. MOM is the standard momentum strategy, the other are dMOM strategies with volatilities calculated using respectively the mean variance (BSC), a Monte Carlo simulation (MC), a GJR-GARCH (GJR), and machine learning (ML). For details regarding variable construction, see Section 3.2

# APPENDIX C: Distributions of Portfolio Returns

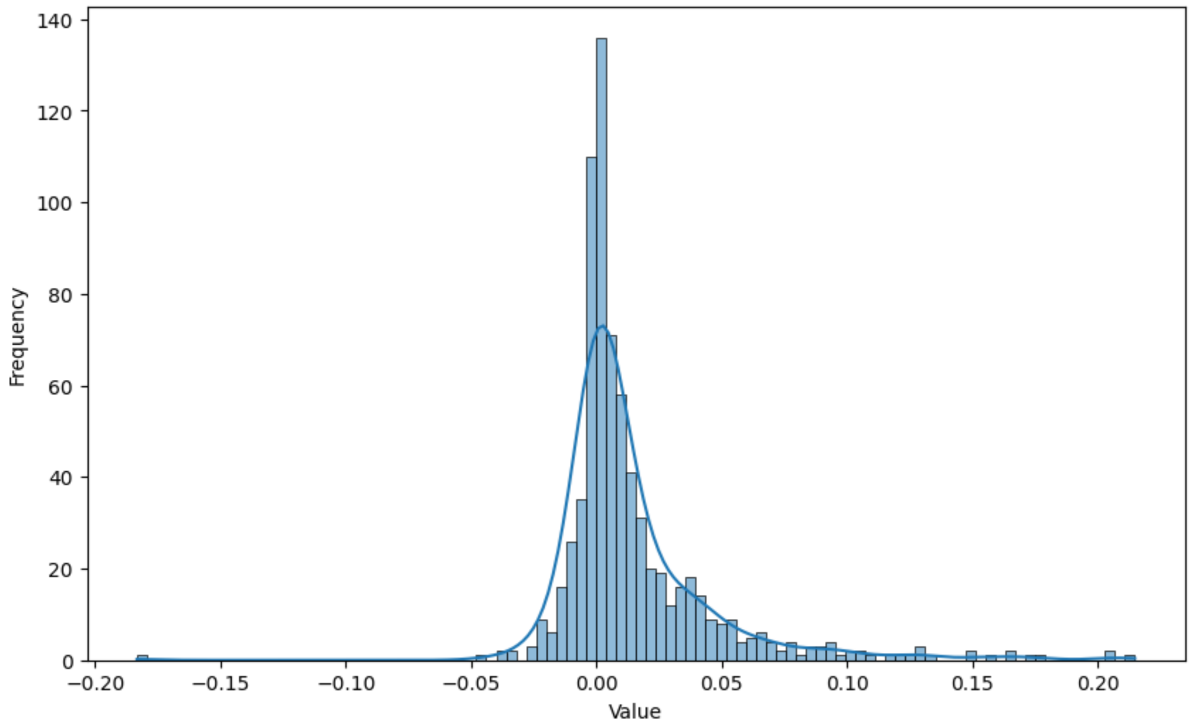


**Figure C1: Distribution of BSC returns.**

This figure displays a histogram, portraying the frequency of each value of returns of BSC over the whole sample, with a distribution curve. The values of "Values" are in USD. For details regarding variable construction, see Section 3.2.

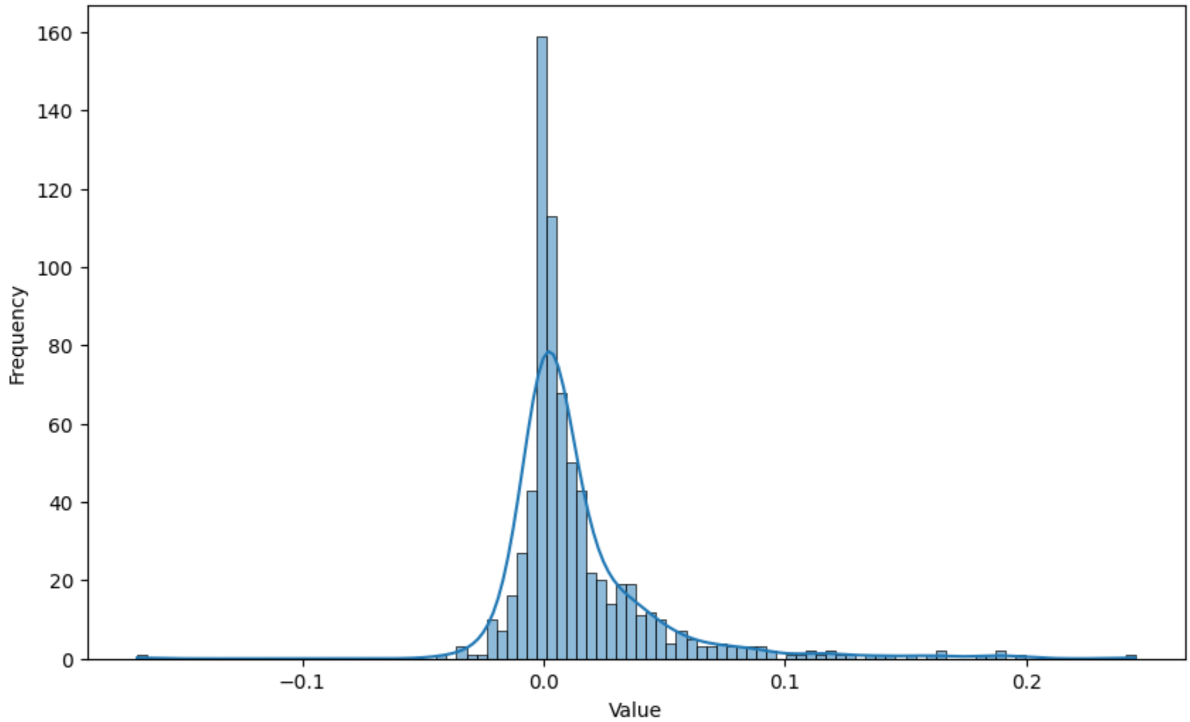


**Figure C2: Distribution of MC returns.**

This figure displays a histogram, portraying the frequency of each value of returns of MC over the whole sample, with a distribution curve. The values of "Values" are in USD. For details regarding variable construction, see Section 3.2.

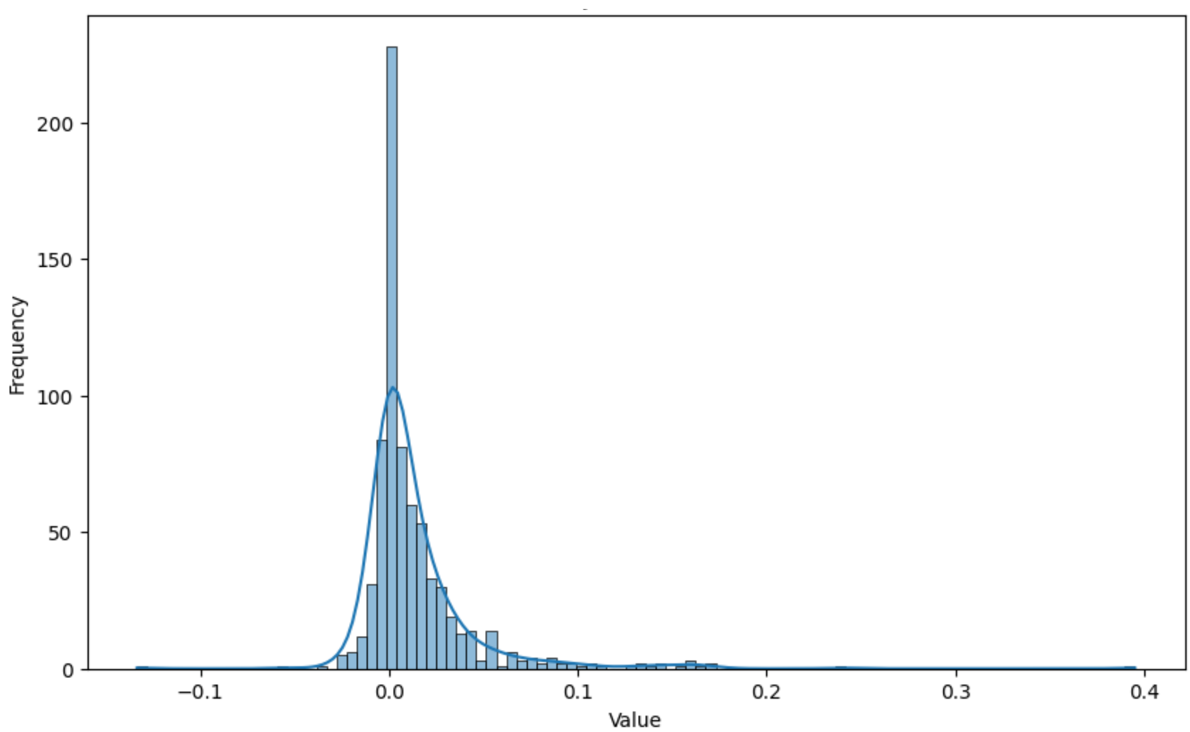


**Figure C3: Distribution of GJR returns.**

This figure displays a histogram, portraying the frequency of each value of returns of GJR over the whole sample, with a distribution curve. The values of "Values" are in USD. For details regarding variable construction, see Section 3.2.

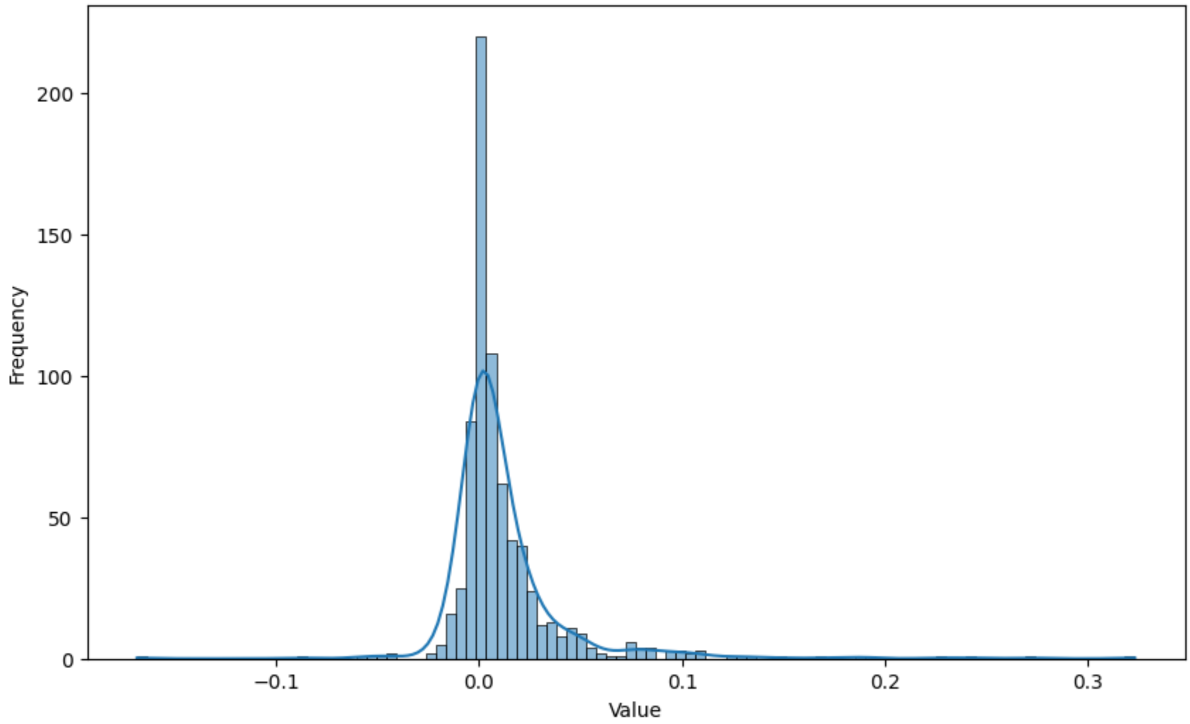


**Figure C4: Distribution of ML returns.**

This figure displays a histogram, portraying the frequency of each value of returns of ML over the whole sample, with a distribution curve. The values of "Values" are in USD. For details regarding variable construction, see Section 3.2.