Trend-Following Strategies for Cryptocurrencies with Machine Learning

Matteo Mamino

Dissertation written under the supervision of Dan H. Tran

Dissertation submitted in partial fulfillment of requirements for the International MSc in Finance, at the Universidade Católica Portuguesa, 04/01/2023.

Abstract

Title: Trend-Following Strategies for Cryptocurrencies with Machine Learning

Author: Matteo Mamino

Cryptocurrencies could bring big returns, but they also carry high volatility and big crash sizes. I discovered that trend-following strategies help investors to mitigate cryptocurrency's risk. I also tested and confirmed that risk managed momentum strategy is applicable to the cryptocurrency environment and that machine learning implementation further improves volatility reduction.

Resumo

Título: Trend-Following Strategies for Cryptocurrencies with Machine Learning

Autor: Matteo Mamino

Cryptomoedas poderão levar a retornos elevados, contudo também podem estar expostos a maior volatilidade e quedas excessivas do mercado. Eu descobri que estratégias que seguem tendências ajudam investidores a reduzir o risco das cryptomoedas. Também testei e confirmei que estratégias que gerem o risco de momentum podem ser aplicadas a cryptomoedas e que machine learning contribui para reduzir a exposição a volatilidade.

Index

1. Introduction

Cryptocurrencies, a new asset class formed in the last decade. A new sector poised to revolutionize the world of traditional finance. Bitcoin, 2009, its inception is the key enabler for the cryptocurrency concept in the financial industry. In the last decade, Bitcoin has been considered both store of value and a valid currency. Its limited supply (fixed to 21 million coins) makes it similar to gold. Furthermore, Bitcoin has a strong track record of stability and appreciation. Its price is certainly volatile in short term, but it has always consistently increased in value over the long term. Meanwhile, an opposite conception has also formed, which sees bitcoin more as a currency. Its capacity to be divisible, meaning that it can be divided into smaller units called Satoshi, facilitates transactions. Finally, the core strength is the decentralization feature, Bitcoin is not controlled by any central authority, such as a bank or a government. This type of structure gives to Bitcoin a level of independence and security that is not matched by traditional currencies. Thanks to this property, Bitcoin is largely used in third-world countries with populations that often face a lack of financial infrastructure. Bitcoin has certainly played a key role in the rise of the cryptocurrency world in common knowledge, and it opened the development of numerous crypto projects. In 13 years, a lot of utilities have been developed making cryptocurrency one of the fast-growing areas of our society. They differ in several ways from the assets best known to us as stocks and commodities. Starting from the fundamental side to the more complex investor behavior. In recent years, trend analysis has been one of the most sought-after and ambitious areas in finance. It is a type of analysis that fits perfectly with the nature and characteristics of the cryptocurrency world. But it's important to address some main differences. The crypto market is well-known volatility and several factors contribute to it. First, limited adoption, cryptocurrencies are still not widely accepted for different use cases which means that their demand is relatively low compared to more established assets. This condition can be the cause of the high market sensitivity to news and sentiment change. Also, limited regulation could represent a key factor for price volatility, an unregulated market is more susceptible to market manipulation and overall unethical behavior. Furthermore, the speculative nature of the market plays a main role too, crypto investors buy and sell exclusively based on the expectation of future price movements rather than on fundamental value. And finally, the limited liquidity of the market makes it more difficult to buy or sell a large amount of cryptocurrencies without significantly affecting their price. Together with volatility, in the recent period, specifically in 2022, the crypto market experienced some extreme events. After an overall and considerable bull market having its peak at the end of 2021, a so-called black swan event took place in May 2022. One of the biggest and most used blockchains, the Terra Ecosystem with its utility token Luna and the stablecoin UST, experienced a major collapse. Its stablecoin broke the peg with the US Dollar, triggering a spiral of events that brought a network worth around $60 billion, to zero. Due to the catastrophic effect on the market, it is estimated that the Luna meltdown caused a loss of around $300 billion in value across the entire crypto environment. Subsequently, some of the most important crypto firms were forced to liquidate or file for bankruptcy, from the crypto lenders Voyager, Celsius, and Blockfi, to the crypto hedge fund 3 Arrows Capital to the last and most unexpected FTX exchange. Keeping in mind that, 2022 in particular, has been one of the worst years for cryptocurrencies in terms of performances and that black swan events are not predictable, it's not a surprise that this new asset class also contains the possibility of great returns and use cases. Cryptocurrencies may provide diversification for an investor's portfolio because they are not directly tied to traditional financial markets or economies. Their value is largely determined by supply and demand, together with a more speculative investors' approach. Another potential benefit of investing in cryptocurrencies is their potential hedge against inflation. Indeed, the supply of many cryptocurrencies is limited and fixed, and in case of high demand, these assets may be less susceptible to inflation than fiat currencies. Some investors could also be attracted by the decentralized and secure nature of blockchain technology, which can be particularly appealing in countries with unstable governments and economies. Finally, cryptocurrencies can also grant access to other types of investments and opportunities, such as Decentralized Finance (DeFi) and Non-Fungible Tokens (NFT). DeFi is a financial ecosystem built on top of decentralized protocols. It is one of the fastest-growing sectors and it is revolutionizing the way of managing digital assets. A DeFi ecosystem grants access to a wide selection of services and products: decentralized lending and borrowing, decentralized asset trading, decentralized derivatives trading, and decentralized insurances. Meanwhile, NFTs represent the gamified side of cryptocurrencies, they are often used as digital collectibles and they are the main pillar of the rising decentralized gaming industry. It is now clear that a high level of risk also brings an equal level of opportunities and possible profits.

The research question lays in this trade-off, is it possible to avoid cryptocurrencies' crashes and volatility with trend-following strategies?

The dissertation author wants to review and shape famous trend-following strategies for cryptocurrency applications with machine learning by contextualizing and leveraging past research to achieve positive and significant results.

2. Literature Review

This paper would contribute to the fast-growing literature on cryptocurrencies.

Strategies carried out by past research in traditional finance could offer insights and inquiry into applicability in the crypto environment. For example, momentum represents one of the best metrics regarding trend analysis with cryptocurrencies. This statement finds confirmation in past literature, it has been shown that momentum strategy performs better with cryptocurrencies that have more attention which are also the most capitalized ones (Liu et al., 2019), meanwhile, other factors like size effect are more concentrated among smallest coins (Liu et al., 2019). It has been shown that momentum not only has a positive effect on cryptocurrency returns but also does not significantly expose it to the possible price shocks to which it exposes stocks (Liu et al., 2019).

Further studies on cryptocurrency confirm that the momentum effect indeed applies to the largest market-capitalized cryptocurrencies, while it causes a short-term reversal with the less capitalized ones. (Cong et al., 2020). Aspects about smaller coins have also been shown by Dhawan & Putniņš (2022) who states that smaller and illiquid coins are often targeted by manipulators to perform pump and dump strategies, they also conclude that these types of activities are attractive for overconfident individuals who are chasing big returns in exchange with extreme risk. This theory is further confirmed by Hackethal et al. (2021) who conclude that cryptocurrency investors have a higher gambling tendency after their first crypto purchase and that they also tend to increase their trading activity.

During the continuous development of the blockchain and cryptocurrency environment in the past years, it has been particularly evident in the impact that communities and user adoption had on the markets. Indeed, most of cryptocurrencies can be identified as a token platform that have a strong relationship with users. Broad research has been conducted on this topic, it’s been demonstrated that platform tokens reduce the volatility of the user base (Cong et al., 2020) probably because users are optimistic for possible upcoming price appreciations (Sockin & Xiong, 2020). But at the same time, a bigger community increases price volatility (Cong et al., 2020), which prevents more widespread use of the platform, this consequence is well explained by the so-called structural amplification of volatility (Pagnotta, 2021). The infamous cryptocurrency volatility could even be explained by the fact that users' and developers' demand could be correlated with investors and speculators. This characteristic explains why cryptocurrency prices may fluctuate wildly (Shams, 2020).

Further research on users showed that the perception of an increase in user adoption by users has an amplified impact on prices (Shams, 2020) meanwhile they tend to under-react to fundamental information (Sockin & Xiong, 2020). More literature deeply researched the fundamental side of crypto assets, a complex factor model has been developed by Bhambhwani et al. (2021) with two factors taken into consideration such as network size and computing power, they showed that these two factors explain crypto returns as good as that of models with more traditional factors such as market, size, and momentum. The two fundamental factors work as a proxy for systematic risk in cryptocurrency markets (Bhambhwani et al., 2019). Additional research conducted by (Cong et al., 2021) propose a 5-factor model that adds two new factors, novel value and network factors to the most traditional factors such as common market, size, and momentum. Finally, also the blockchain baseline has been assessed by some literature on miners' relationships and decentralization. Pagnotta & Buraschi (2018) state that an increase in the number of miners for PoW (Proof of Work) causes and higher price as a consequence of the intense competition and increased hash rate, but at the same time, the increase of the network causes network delays consequently prolonging the consensus generation (Hinzen et al., 2022). It has also been demonstrated that decentralization, which is the main strong point of blockchains, could also cause coordination failure and inefficiency in PoW blockchains such as Bitcoin (Biais et al., 2019). Basically, the miners tend to individually choose their optimal computing capacity, causing a negative effect for other miners through increasing difficulty (Biais et al., 2019). Four out of the nine variables included in the research are Bitcoin-based features. Bitcoin's role in the cryptocurrency market has been one of the most studied branches in past literature. Narayan et al. (2016) is a comprehensive research on Bitcoin and its relationship with other cryptocurrencies. The study highlights how Bitcoin is often seen and perceived as a front-runner for the crypto industry. A more quantitative approach has been conducted by Kalenova et al. (2018), the study found that there is a strong positive correlation between Bitcoin and the overall crypto market. The positive correlation has also been sustained by Hou et al. (2019), the research shows that the relationship between Bitcoin and the rest of the market is complex and not always straightforward, but they conclude and confirms that Bitcoin generally has a strong influence on the overall market.

As mentioned before, momentum is certainly an interesting topic both in cryptocurrency and in the traditional finance world. A lot of past studies have shown different points of view on momentum research, from its correlation to the value factor explained by Asness et al. (2013) to the possible nature of momentum itself discussed by Conrad and Kaul (1998) who state that momentum profits come from the cross-sectional difference in expected return rather than time series returns pattern. There have also been studies regarding behavioral patterns that could explain momentum Hong & Stein (1999). An innovative approach to momentum has been shown by Barroso & Santa-Clara (2015), their research shows that the risk of momentum it's highly predictable and can be managed, eliminating the risk of crashes, and increasing the Sharpe ratio. Such an approach is highly innovative and paves the way for countless new research. A close and careful look at past machine learning studies was fundamental and set the basics for the writing of this research. Starting from technical studies like data preprocessing approaches analyzed by Xing (2019) who states that feature scaling improves the algorithm performances and that normalization and standardization speed up the iterations together with improving efficiency. Christensen et al. (2021) discovered that random forests and artificial neural network algorithms are preferable for volatility forecasting. They also stated that the models are also robust to noise variables. Mittnik et al. (2015) suggest that also gradient boosting approach conducts significant results with volatility forecasting, they also added that financial variables have a high non-linear influence on volatility.

3. Data

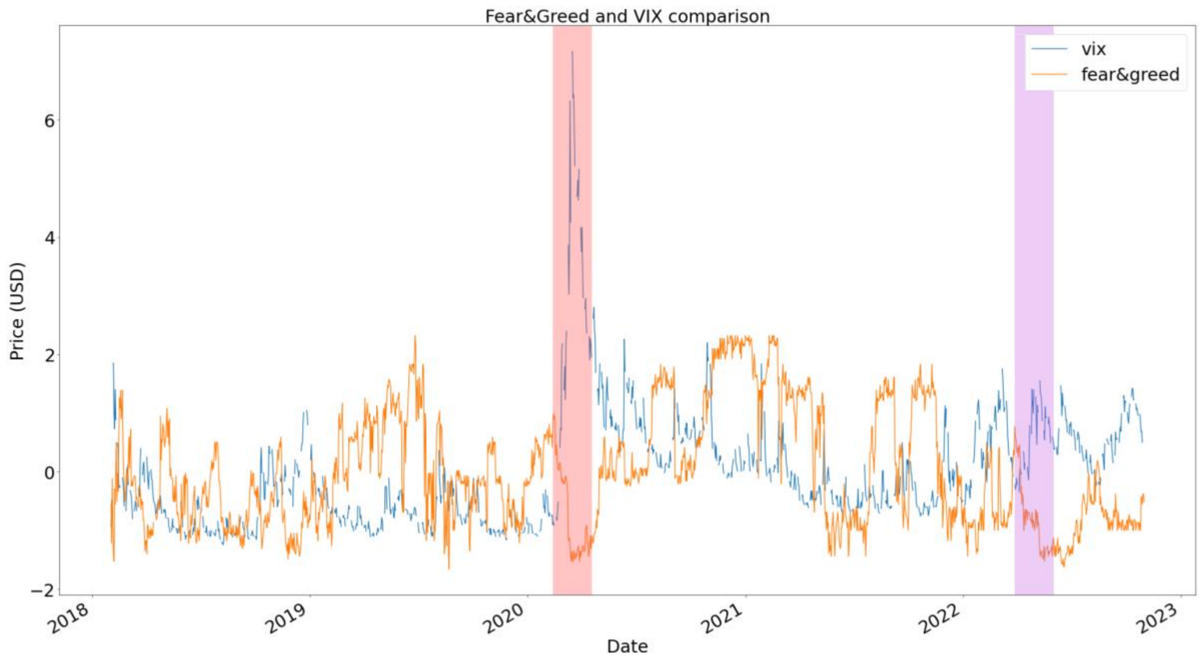
To conduct the research, I accessed three different types of data: cryptocurrency-related data, macroeconomic data, and technical indicators.

3.1 Cryptocurrencies features

Historical prices of the top 500 cryptocurrencies for market capitalization from Yahoo Finance, I would like to emphasize that the provider of the raw data is CoinMarketCap. Again, from CoinMarketCap I downloaded the market capitalization of the top 500 cryptocurrencies to conduct some analysis on size further down the research.
As suggested by Shams (2020) a perception of an increase in user adoption has positive effects on cryptocurrency prices, accordingly, I used the Bitcoin daily volumes as a proxy for user adoption. To check and exploit the effect described by Pagnotta & Buraschi (2018) I also retrieved the historical daily data of the Bitcoin hash rate. The hash rate is the measurement of the computational power of PoW (Proof of Work) blockchains like Bitcoin. A hash is a mathematical function that takes an input and produces a fixed output. In the case of Bitcoin, the inputs are blocks of transaction data. The hash rate is important because it determines how quickly new blocks are added to the blockchain. In order for a new block to be produced, a group of computers must solve a complex mathematical problem and produce a valid hash value for the block. If the hash rate is high, more computers are working on the problem (higher miners' competition), and the probability of finding a fast solution increases. If the hash rate is low, it takes longer to find a solution, and the new blocks are added at a slower pace. Indeed Pagnotta & Buraschi (2018) state that the increased hash rate and miners' competition have a positive consequence on Bitcoin price.

3.2 Sentiment features

I retrieved the historical values of the Fear and Greed Index, which is a tool used to measure the emotional state of the cryptocurrency market. It is based on a range of factors, including market volatility, volume, social media activity, and the dominance of certain cryptocurrencies. The index uses a scale of 0 to 100, with 0 representing extreme fear and 100 representing extreme greed. When the index is low, it may indicate that investors are feeling fearful or uncertain about the market and may be more likely to sell their assets. On the other hand, when the index is high, it may indicate that investors are feeling optimistic and greedy, and may be more likely to buy assets. The fear and greed index is not a definitive indicator of market direction, but it can be useful for understanding the sentiment of the market and for identifying potential buying or selling opportunities.



= Covid-19 = Luna ecosystem crash

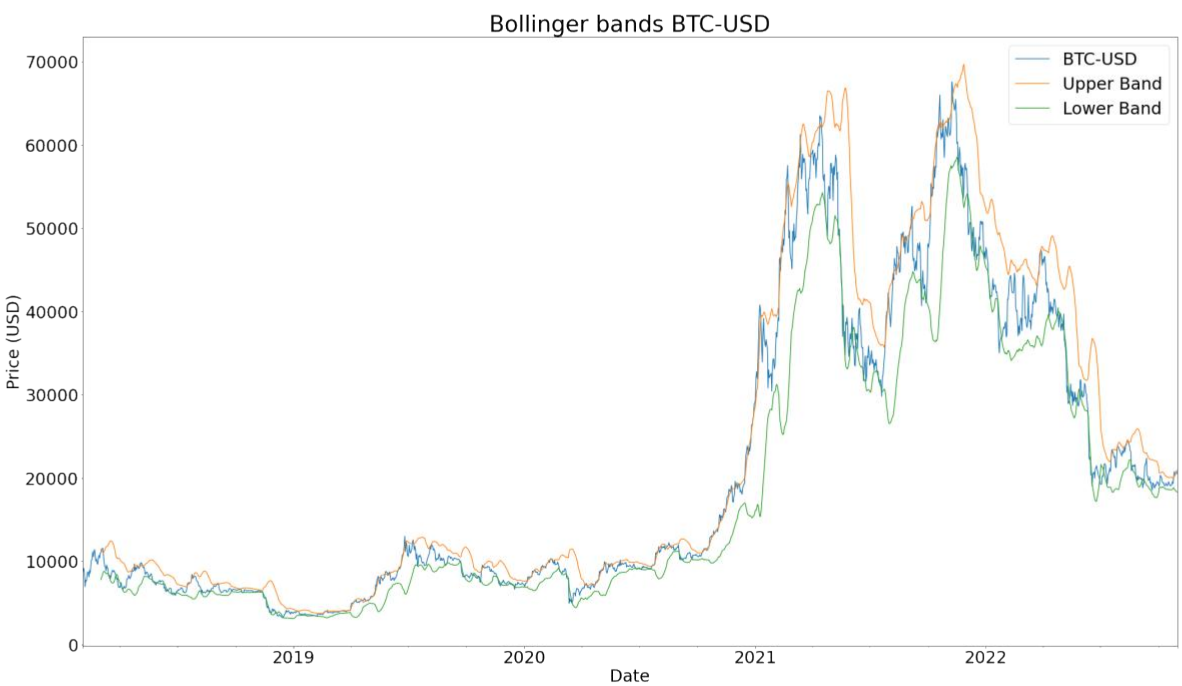
A direct example of a "fear index" in the traditional financial market can be the VIX, because it tends to rise when volatility increases and investors are more fearful of potential losses. In the graph, we can see how both sentiment indexes reacted to a big systematic event like the Covid-19 outbreak, with a big spike in VIX level and a big crash in the Fear and Greed score.

3.3 Macroeconomics features

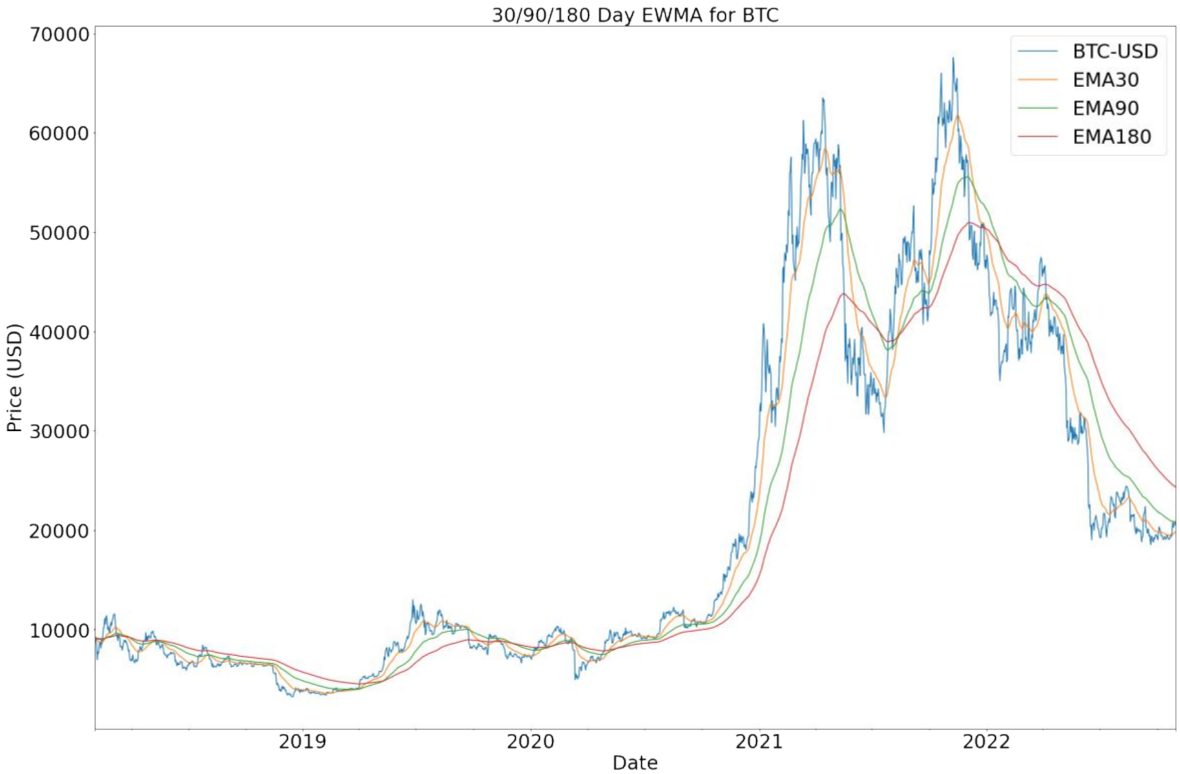
The macroeconomic indicators that I included are S&P500 daily returns and VIX daily changes. The inclusion of such variables is suggested by different past research that states that there is a negative correlation between VIX and cryptocurrencies' volatility. Different authors suggest that such behavior maybe be caused by investors seeking out alternative investments, such as cryptocurrencies during times of market instability. These suggestions come together with the specification that the negative correlation is significant just on some crypto assets rather than the entire market. A negative correlation has also been found between S&P500 returns and crypto volatility. Over time the cryptocurrency market would likely influence and react to the traditional financial system in a more clear and linear way. Currently, the cryptocurrency market is still an immature and less liquid market, with less attention from retail and institutional investors.

3.4 Technical features

Technical indicators are based on historical data and are particularly helpful for trend-following purposes. I focused on indicators with a high volatility forecasting power. Bollinger Bands consist of three lines, a moving average in the middle and an upper and lower band. The upper and lower band are usually set at two standard deviations above and below the moving average. The idea behind the Bollinger bands is that the price tends to stay between the upper and lower bands, but it can break out as a result of trend change or volatility. If the price breaks out of the upper band, it's considered a sign that the market is overbought and that the price could potentially fall. Similarly, if the price breaks out of the lower band, it may be a sign that the market is oversold and that the price could potentially rise. The Bollinger bands are commonly used to identify trends and to measure the volatility of the asset in consideration.



Among others, Barroso & Santa-Clara (2015) state that exponentially weighted moving averages can be used to forecast the variance of returns, so I proceeded with the inclusion of one, three, and six months EWMA. Exponentially weighted moving average (EWMA) is a technical indicator that gives more weight to recent observations. It’s mainly used to identify long-term trends, remove noisy data, and to highlight underlying patterns.



The dataset’s timeframes are different from each other, the cryptocurrencies’ historical daily prices range from 2014 to October 2022, while the fear and greed index from January 2018 to October 2022, and finally the historical daily Hash rate ranges from 2009 to October 2022. For research and comparison purposes all the timespans of the datasets were unified with the following characteristics, beginning 01/02/2018 and ending 31/10/2022, bringing the total observable days to 1735.

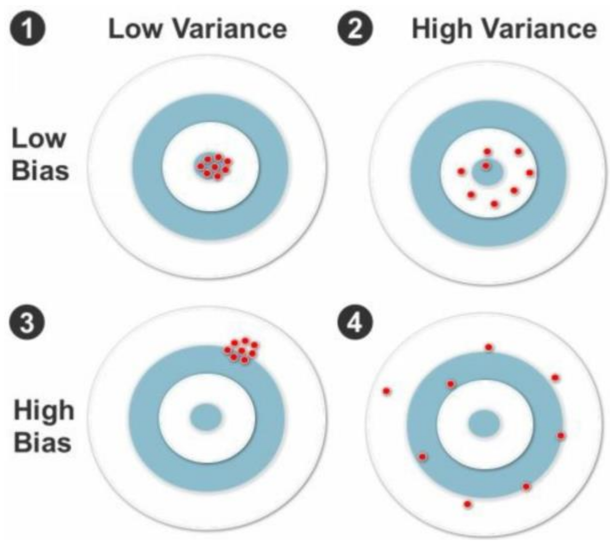
4. Machine Learning

Machine learning was first conceptualized by Samuel (1956) as a subfield of computer science. It was considered as the computer capacity of learning without user intervention also called 'self-learning'. Based on data and empirical information, it relates to the use of statistical approaches to identify trends and enhance performance. Basically, the machine is simulating the process of human-based decision-making by creating decisions based on experience. Machine learning can be visualized as a three-step process: Data-Model-Action. Data is the base of the entire machine-learning workflow. First, it's fundamental to get data from official and trustworthy providers. Second, it's then possible to clean, transform and shape the datasets based on the need of the pursued strategy. The process which transforms the raw data into prepared data is called data engineering and It's constituted by data cleaning and transformation processes. Thanks to this step, it's possible to conduct research and find helpful insights regarding possible relationships and influences inside the dataset. The most common data cleaning operations are removing/replacing missing values, filtering outliers, remove unwanted and duplicated information. Once the dataset is clean and well structured, the feature engineering process takes place, preparing the dataset for the machine learning model fitting. The most common feature engineering operations are feature splitting, variable transformation, categorical encoding, and data scaling. Finally, the dataset has been cleaned and customized according to the requested criteria and it's now ready for the train/test split. In this fundamental step, the dataset is split in two, usually with a ratio of 80:20. The first part is the train set which represents all the data that will be used to develop and train the model. The second part is the test set which will be necessary to test the performances of the model.

The train/test split section potentially contains one of the biggest flaws in a machine learning approach. If the data split has not been performed carefully, data leakage occurs. Data leakage is the situation in which the model is trained using part of the values of the test set. In this condition, the prediction based on the train set and part of the test set would be biased, and it would cause an invalid model prediction. The biggest problem of data leakage is that it can go unnoticed, resulting in an overestimation of the prediction. To avoid data leakage, it's important to cut the last values of the train set and the first values of the test set.

Noticing underfitting and overfitting, which characterize how closely your model matches the real patterns of the dataset, is a serious problem in machine learning. Understanding bias and variance is a prerequisite for comprehending underfitting and overfitting.

Bias/Variance trade-off



Usually, high bias translates to an oversimplified model that leads to a high error on both training and test data, causing underfitting. On the other side, high variance consists of an almost perfect fit for training data but with high error on test data, causing overfitting. The bias/variance tradeoff is usually managed with a change from linear to non-linear algorithms, hyperparameters tuning, or feature addition/remotion. Numerous statistical-based algorithms are used in machine learning, and selecting the best algorithm is one of the biggest challenges. Before diving deep into the models used in this research it's important to understand the difference between the three main types of machine learning models.

Supervised learning works using labeled data as inputs. The labeled data, as previously stated is divided into features "X" and the value to forecast "Y". The machine works to identify patterns and relations between the variables and then it tries to recreate them, making a forecast of "Y". To provide further understanding, the supervised learning algorithm knows the value of "Y" and it works backward to determine the influences between the "X" features and "Y". Finance cases usually imply a supervised learning algorithm due to their organized and linear nature.

The best-known supervised learning models are regression analysis, supported vector machines, neural networks, and decision trees.

Unsupervised learning works with the same logic as a supervised learning algorithm but with unlabeled variables, the biggest advantage is that the algorithm identifies relations that are unnoticed by the user. Unsupervised learning is particularly useful for fraud detection.

The most used unsupervised machine learning models are k-mean clustering and neural networks.

Reinforcement learning can be considered the most complex and advanced approach. It works by collecting inputs from previous iterations and constantly refining the model.

Next, the three machine learning models used to forecast the realized variance of the momentum strategy are : Artificial Neural Network (ANN), Gradient Boosting, and Random Forest. They have been selected based on past literature (see Christensen et al.,2021 and Mittnik et al.,2015) and on their natural predisposition to regression-based problems.

4.1 Artificial Neural Network

The Artificial Neural Network allows data processing through different analysis layers.

Between the input and the output, there are the hidden layers which are composed of processing units called nodes. Each node receives input from other nodes or from external sources and it uses these inputs to perform a computation to produce an output. The computation performed by a node is based on its weight and biases, which determine how much influence each input has on the output of the node. In respect of neurons in the human brain, the nodes of an artificial neural network are basically simple mathematical functions that operate on vectors of numbers.

The version used is a Multi-Layer Perceptron Regressor (MLPRegressor) due to the variance forecast needed. This type of ANN is particularly effective for learning non-linear models, with the downside of being sensitive to feature scaling. The model has been tuned controlling two different parameters: the learning rate and the alpha, which, by penalizing weights with large magnitudes, aids in preventing overfitting.

4.2 Gradient Boosting

The Gradient Boosting algorithm is one the best bias error minimizer, it can be used for both the prediction of continuous variables (regressor) and categorical variables (classifier). It works by building a series of weak models, also known as "boosters", which are then combined together to form a stronger model. The algorithm is able to improve predictions by focusing on the areas where the previous "booster" model performed poorly and by correcting those errors. The Gradient Boosting Regressor has been used in the research, this version uses a mean squared error (MSE) cost function. The algorithm is suitable for finding non-linear relationships, it can also deal with flaws deriving from poorly processed data such as missing values and outliers. It's particularly flexible and it can be adjusted for a variety of use cases.

4.3 Random Forest

The Random Forest algorithm works by creating multiple decision trees that are then merged in order to generate a more accurate prediction. The strength of the model lays in the difference between a single decision tree and a decision forest. In the case of a decision forest, the single decision trees are randomly put together and they organically grow instead of individually. This process helps to identify non-linear relationships in an easier way. The Random Forest Regressor has been used in this research, the model has been tuned by the number of trees in the forest (number of estimators), by the bootstrap settings and by maximum depth. 60 estimators have been selected together with the use of the bootstrap approach, in this way, random row and features sampling are performed to form sample datasets for each model.

5. Methodology

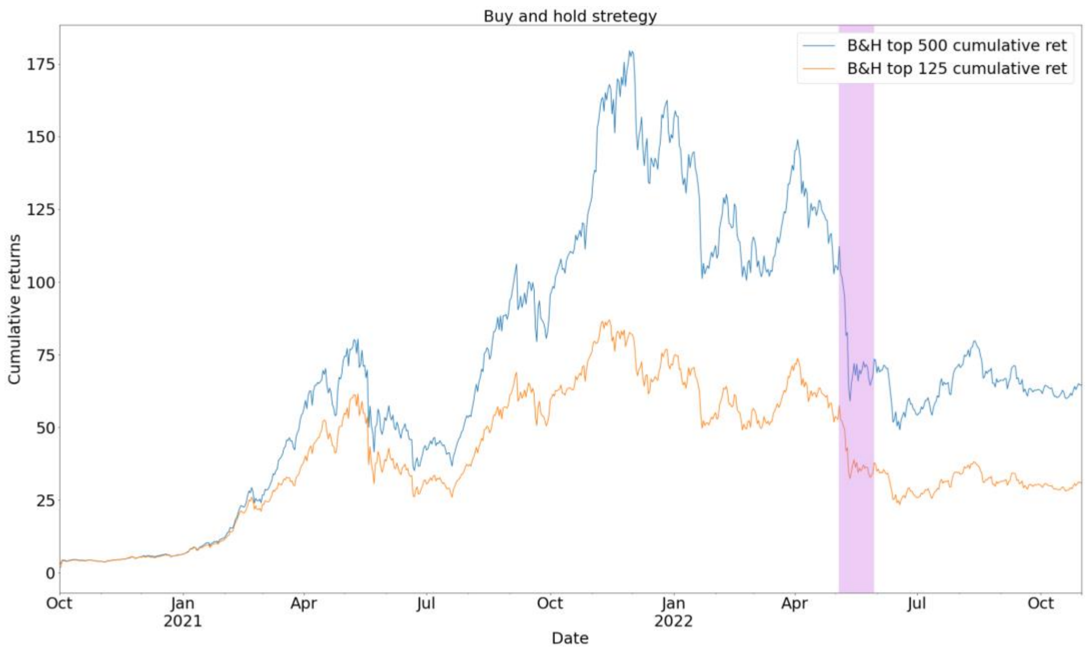
This chapter examines the procedures and actual trend-following strategies applied to cryptocurrency investing. But first, it's fundamental to specify and showcase which type of data processing has been performed. Starting from cryptocurrencies-based processing, 29 stablecoins have been removed from the initial 500 cryptocurrencies dataset, bringing the overall number down to 471. Stablecoins are a type of token that mimics the value of an underlying asset, usually a fiat currency such as EUR or USD or even a commodity like gold in the case of "Pax Gold". Consequently, the cryptocurrencies with less than 6 months of life have been removed, they carry extreme volatility, and noise and are often targets of pump-and-dump schemes (Dhawan & Putniņš, 2022). Outliers have been removed, and missing values have been replaced. The historical daily prices timespan has been reduced, from the original 2014/2022 dataset to a new 2018/2022 dataset. This cut has been done with the intent of reducing the weight of Bitcoin's influence in the period before 2018. Indeed, the average of actively traded cryptocurrencies from 2014 to 2018 inside the dataset is 7, against the 272 actively traded cryptocurrencies from 2018 to 2022. The timeframe shortening has also allowed a more significant comparison with the other sentiment and macro features.

5.1 *Vanilla strategies*

The first step is constituted by the composition of basic and easily replicable crypto trading strategies: buy and hold strategy, trend following with technical indicators (Simple Moving Average), with sentiment score (Fear & Greed Index), and classic momentum strategy. This section has been conceptualized as a basket of strategies that are accessible and used by many investors, and it can be considered as an introduction to trend following for cryptocurrencies.

a. Buy and hold strategy

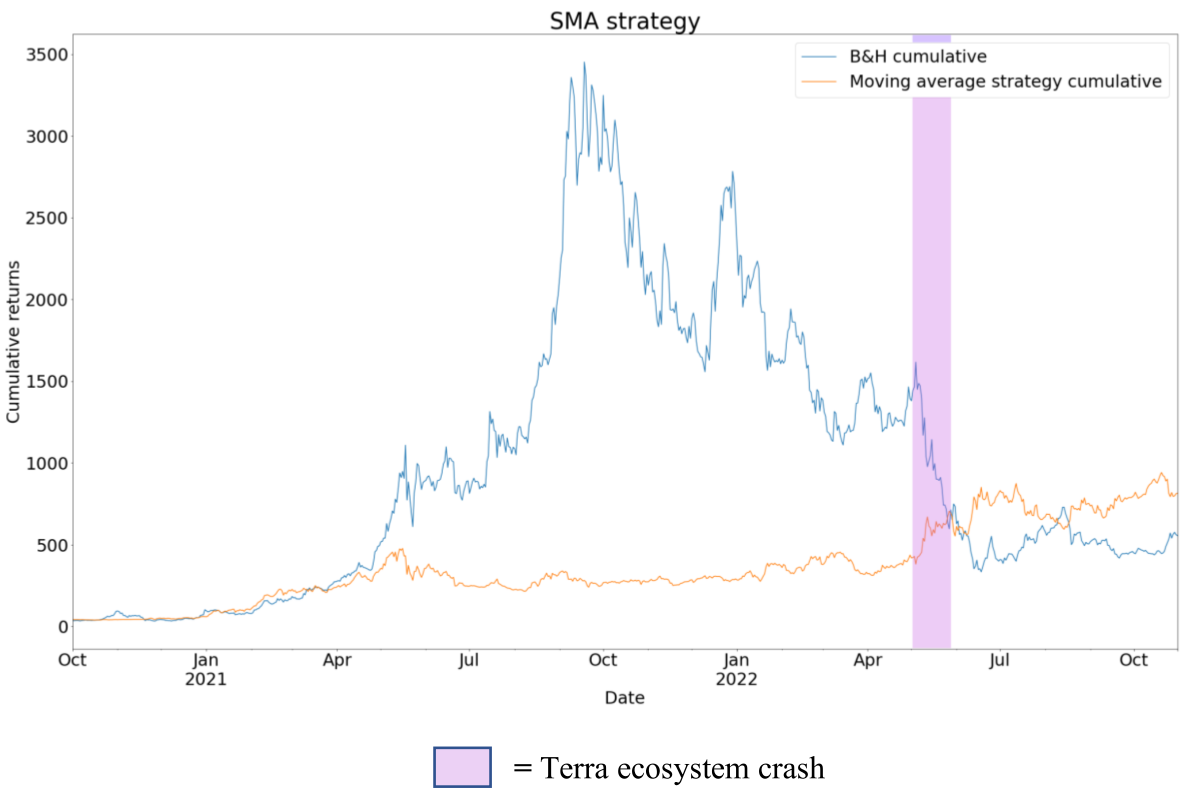
This passive strategy is one of the favorites of retail investors, it consists in buying the assets and holding them in the portfolio waiting for long-term appreciation. I included it to have a basic and reliable benchmark strategy. I performed it with three different datasets, the top 10 cryptocurrencies by market cap, the top 125 cryptocurrencies by market cap, and the top 500 cryptocurrencies by market cap. The buy-and-hold strategy has been selected as the benchmark to back-test the effects and performances of the other approaches.



= Terra ecosystem crash

b. Trading strategy with moving averages

Trading using moving averages is one of the simplest and most used technical trading strategies. Moving averages help to identify trends in asset prices. If Bitcoin's price is consistently above its moving average, it may be in an uptrend. On the opposite side, if Bitcoin price is consistently under the moving average, it could be in a downtrend. In addition, moving averages can be used to identify potential support and resistance levels, but I did not include this particular application in the current strategy. The concept behind this strategy it’s simple and straightforward. The strategy consists of a dynamic adjustment of the portfolio based on the relation between a long and a short moving average, specifically in the situation where the two moving averages cross each other. This can be either a bullish or bearish signal. If the short moving average (SMA) crosses above the long moving average (LMA), it indicates that the trend is shifting upwards, and it may be interpreted as a buying opportunity. In the opposite case, if the SMA crosses the LMA, the situation is interpreted as a selling opportunity. The shorter the moving average period, the more sensitive it is to price changes, while the LMA is more representative of the whole trend. The strategy is applied to the top 10 cryptocurrencies by market capitalization. The long moving average is based on past 150 days while the short moving average on the past 5 days. The core idea is to go long when the short-term MA is bigger than the long-term MA and go short when the long MA is bigger than the short MA. The 5 days MA, with such short timeframe, is sensible to current price volatility and it identifies possible trend inversions.



c. Trading strategy with sentiment index

A trading strategy based on a sentiment index could be successful because it allows the investor to grasp the overall sentiment of the market or of a specific sector. A sentiment index could result in being a useful tool because it helps to identify trends that may not be immediately obvious from analyzing traditional financial metrics. A sentiment index may show that investors' sentiment towards a particular sector is becoming increasingly negative even if the underlying fundamental metrics appear to be strong. Furthermore, trading using a sentiment index can give a sense of how the market is going to perform in the short term. This aspect is particularly useful for the volatility and crash forecasting goals of the current research. Overall, a sentiment index such as the Fear & Greed Index could be useful for investors who are looking to make strategic moves based on Bitcoin sentiment conditions. The index taken into consideration for this strategy is the Fear & Greed Index. It reflects the daily sentiment level of Bitcoin with a score that ranges between 0 to 100. The higher the score, the higher the level of greediness in the market, the lower the score, the higher the level of fear in the market. The score is composed and influenced by different factors such as Bitcoin dominance, Bitcoin volatility, social media sentiment, and Bitcoin volumes. The dataset used is represented by the top 125 cryptocurrencies by market capitalization (top 25%). The simple rule of the strategy consists in going short with the whole portfolio when the previous day registered a fear score smaller than 20 and going long with the entire portfolio in the opposite case. The concept is to capture the start of a period of fear and crashes with a fear score smaller than 20. The value has not been chosen arbitrarily, the average of the Fear & Greed Index from 2018 to 2022 is 42 with a standard deviation of 22. It means that values under 20 can be considered periods of extreme fear and uncertainties.

= Terra ecosystem crash

d. Momentum strategy

A classic momentum strategy has been performed, trying to exploit both winners' and losers' trends. The core idea behind a momentum strategy is the premise that assets that have been performing well in recent past periods are likely to continue to do so in the near future. This is because momentum often reflects underlying trends in the market or in a particular industry, and these trends can persist for extended periods of time. A momentum strategy could also help an investor to identify emerging patterns. This can be particularly useful for investors who are looking for portfolio diversification because it may help to navigate emerging industries or sectors. For this strategy, the dataset used comprehends the top 125 cryptocurrencies by market capitalization, based on the discovery that momentum has a positive effect only on large-cap cryptocurrencies (Cong et al., 2021). A 10-day rolling window has been used to compute assets' cumulative returns, skipping the most recent day according to Jegadeesh and Titman (1993) and Asness (1994). Further on, I created 9 percentiles for each day, percentile 10 being the one with the lowest values and percentile 90 being the one with the highest values. Then, I distributed the 10 days cumulative returns for each asset in 10 portfolios, sorting them using the percentiles previously created. Portfolio 1 represents the average returns of the bottom 10% losers, while portfolio 10 represents the average returns of the top 10% winners. As the final part of the strategy, I proceeded to short all the components of portfolio 1 and buy all the components of portfolio 10 every day.

Some descriptive statistics have been computed in order to directly check the efficiency and performance of the strategies.

Vanilla strategy results

	B&H	SMA	F&Gindex	ClassicWML
Mean (%)	89,17%	36,10%	79,58%	126,29%
Std. Dev. (%)	79,28%	62,98%	83,53%	79,94%
Sharpe ratio	1,1247	0,5733	0,9526	1,5798
Skewness	2,1810	0,0379	6,0189	0,9129
Kurtosis	42,3050	7,7468	138,3424	8,9048

The classic momentum strategy with the top 125 cryptocurrencies outperformed the others with a higher Sharpe ratio. The Sharpe ratio indicates the reward for every unit of downside risk and it’s a key indicator of a strategy's efficiency. Also, skewness and kurtosis represent fundamental indicators. The skewness, if it is positive, it communicates that the distribution presents few losses and few large gains while if it’s negative, the opposite situation occurred. The kurtosis metrics indicate the degree of possibility of tail events. The higher the kurtosis, the higher the number of tail events inside the distribution. Overall, the strategy with less volatility is the one exploiting the relationship between the short and long moving averages, with a final value of 62,98%.

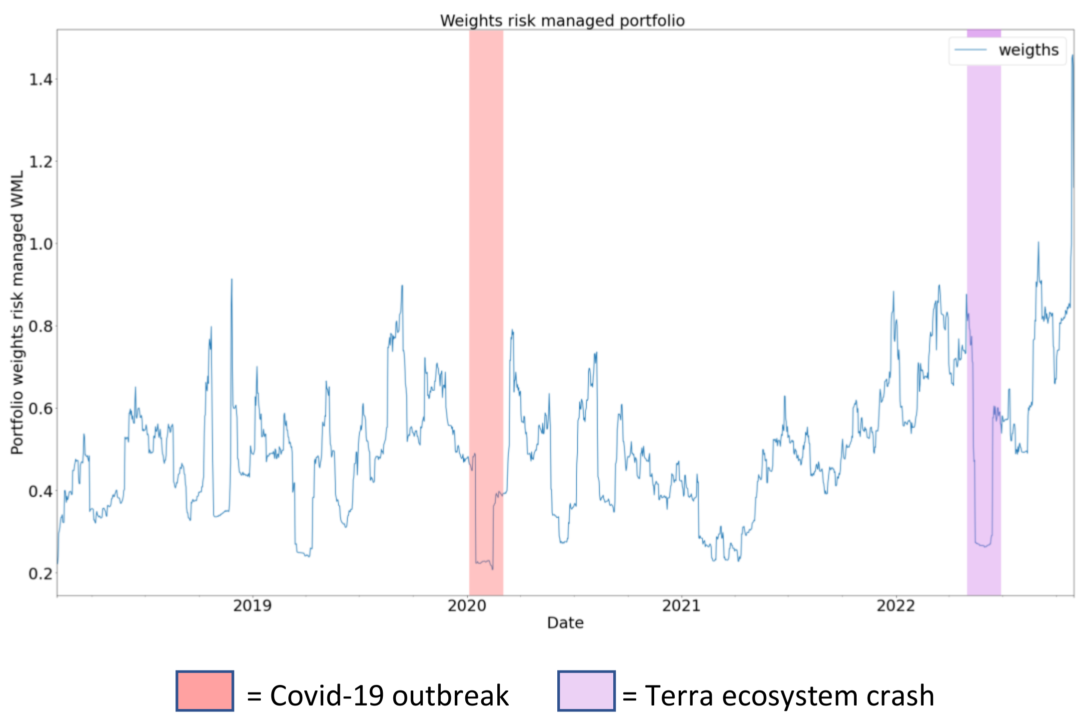
5.2 Risk managed momentum

The second step constitutes an improvement of one of the strategies that are part of step one. Step two is intended to represent the next level from the basic and simple strategies explained above. The strategy I have performed in this part is the risk managed momentum. Barroso & Santa-Clara (2015) state that risk managed momentum has a great effect on portfolio performances, reducing the crash risk and almost doubling the Sharpe ratio. The strategy is built upon the classic momentum strategy previously explained. I computed the variance forecast summing 30 previous days of the squared returns of the classic WML (winners minus losers) strategy. I reduced the rolling window dimension from the original 180 days to just 30 days because of the already short timeframe.

$$\hat{\sigma}^2_{WML,t} = \sum_{j=0}^{30} r^2_{WML,d_{t-1}-j}$$

Then I computed the weights of the scaled momentum, dividing a target volatility constant of 12% (see Barroso & Santa-Clara, 2015) by the forecasted volatility.

$$weight = \frac{\sigma_{target}}{\hat{\sigma}_t}$$



It's interesting to note how well the strategy reacts to high volatility periods. The weights of the scaled momentum almost went to zero during the start of the Covid-19 pandemic (02/2020) and during the Luna / Terra ecosystem crash (05/2022). The weights of the scaled momentum have a crucial role in the strategy, they allow to scale down the exposition to momentum during the periods of high volatility, which are the ones causing major downsides in a classic momentum strategy. Consequently, the daily weight of scaled momentum is multiplied by the daily returns of the momentum strategy, resulting in the returns of the risk managed momentum strategy.

Let $r_{WML*,t}$ be the returns of the risk managed momentum strategy and $r_{WML,t}$ returns of the classic momentum strategy.

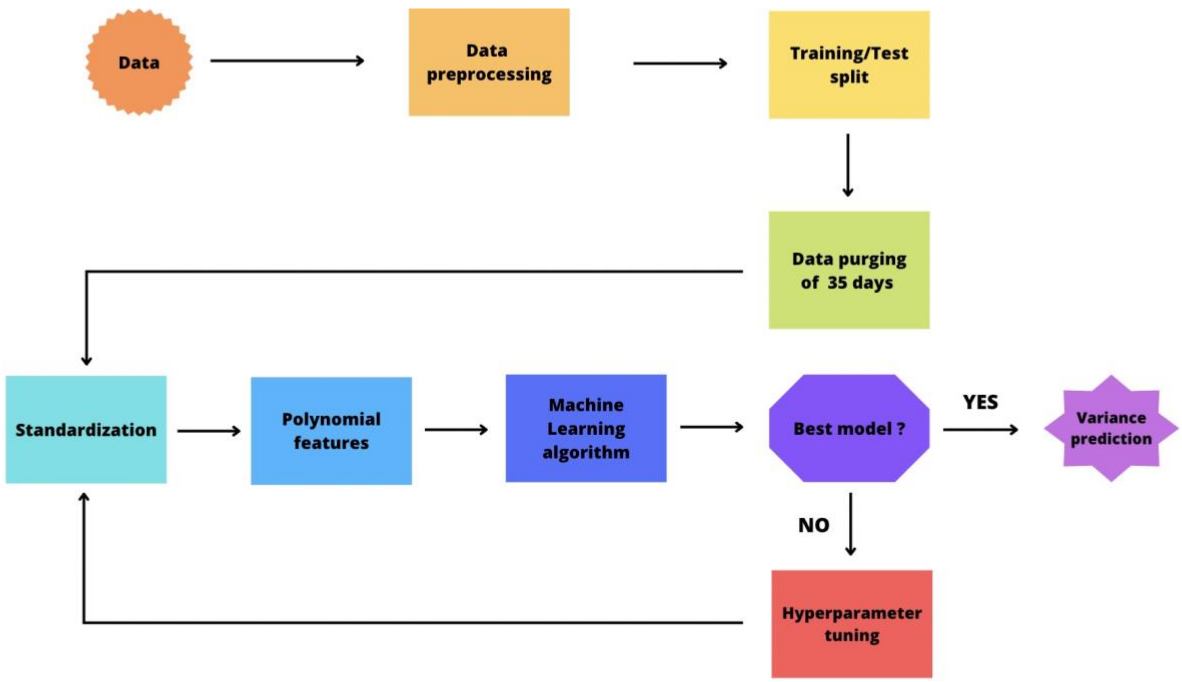
$$r_{WML*,t} = \frac{\sigma_{target}}{\hat{\sigma}_t} r_{WML,t}$$

5.3 *Machine learning implementation*

The third and last step represents the optimization of the risk managed momentum strategy, explained in the second step, with the use of machine learning algorithms. As previously explained in the machine learning section, 3 different algorithms have been performed, Artificial Neural Network, Gradient Boosting and Random Forest. First, data processing is needed before fitting the model. All the "X" variables have been transformed into percentage change; it means that each variable is now showcasing its percentage change from the previous day. This step helps to highlight patterns and to put all the variables at the same format. Subsequently, I lagged six of the thirteen "X" variables using 30 days rolling window. This addition brings the total number of "X" variables to 194. The "X" features lagged are the percentage change of the Fear & Greed Index, the Bitcoin returns, the Bitcoin volatility and the 30/90/180 days percentage change of the exponentially weighted moving average (EWMA). The variables have been selected and lagged due to their expected gradually manifested causal effect on the response variable "Y".
The goal is to forecast the variance of the momentum strategy in order to scale down the exposure before crashes occur. Since the problem has a forecasting nature, all three algorithms have been used with their regressor versions. All three models have been built using pipelines. The process of creating a machine learning model may be codified and automated using a machine learning pipeline. Machine learning pipelines are made up of several sequential processes that handle

everything from model training and implementation to data extraction and preparation. Inside the pipeline I decided to allocate 3 main steps: standardization, polynomial featuring and finally the machine learning algorithm required.

Machine learning workflow



The figure is showing the machine learning workflow used in this research, from the data preprocessing to the algorithm implementation and hyperparameter tuning.

a. Data processing

Preparing the data is the starting point for a machine learning implementation because it helps to ensure that the data in the model is clean, consistent, and in a format that the model can effectively learn from. It is a crucial point to the success of a machine learning model, as the quality of the input data will directly affect the accuracy and performance of the model. Data processing also helps to define the train and test sets that are the fundamental part of a machine learning model. It further helps to extract relevant features and to remove noise and inconsistencies. Without proper data processing, the machine learning model may not be able to learn effectively and may not perform well on data. In conclusion, machine learning models can be only as good as the features they are trained and tested on. Keeping in mind the previous premises, the dataset has been divided into train and test sets with an 80:20 ratio. In order to avoid data leakage, a data purge has been performed. The last 35 days of the train set and the first 35 days of the test set have been removed. In this way, as previously explained, the probability of having a part of the test set inside the training process is avoided. Consequently, the train and test sets experienced two main steps in data processing, standardization, and polynomial features. Standardization is a common technique that is used to transform variables so that they have zero mean and a standard deviation of one. This is done to bring all the variables to the same scale because variables that are on a different scale don't equally contribute to the model. Standardized variables are also easier to interpret and are free of noise effects. The noise reduction takes place because the standardization process minimizes the differences between the variables, this further allows the highlight of underlying patterns in the data. On the other side, polynomial features are a feature transformation that brings input features to a higher degree of polynomial terms. It is mainly used to capture the complex relationship between the input features and the target variable. Using polynomial features can improve the accuracy of the machine learning model, especially when the relationship between the "X" variables and the "Y" variable is non-linear. The use of polynomial features has also a downturn, it can increase the complexity of the model, consequently causing overfitting.

b. Hyperparameter tuning

After processing the data inside the pipeline and executing the first two steps of standardization and polynomial features, the data are ready for the algorithm implementation. The algorithms in use have been downloaded from Scikit-Learn, which is one of the best APIs to use for the creation of machine learning applications. The algorithms have been tuned based on the need of the current use case. Finding the optimal sets of hyperparameters for a particular problem can be challenging. It is the process of systematically searching for the best combination for a given model and dataset. Even small changes to hyperparameters can result in significant changes in model performance.

The Artificial Neural Network has been tuned controlling two parameters: the alpha and the learning rate. The alpha represents the level of strength of L2 regularization parameters which aim to penalize the weights with larger magnitudes, forcing them to smaller values. Regularization is particularly effective in avoiding overfitting by adding bias to the model. The Random Forest algorithm has been tuned controlling three parameters: the number of estimators, the max depth, and the bootstrap settings. The max depth hyperparameter is one of the best to use for random forest tuning because it controls the level of complexity of the model and it's effective against overfitting. Without controlling the max depth, the nodes are expanded until the perfect fit has not been found, exponentially increasing complexity. Meanwhile, the number of estimators hyperparameter controls the number of trees in the forest. If we want a better visualization of how tuning a random forest looks like, we need to imagine the number of estimators as the number of trees in the forest and the max depth as the number of leaves on each tree. The artificial neural network and random forest algorithms have been subjected to regularization due to their nature to create complex models and to cause overfitting.

6. Results

6.1 *Models results*

Machine learning models performance

	Train		Test
	RMSE	R2	R2
Neural Network	0,16927	47,69%	-83,55%
Gradient Boosting	0,06019	87,12%	-33,34%
Random Forest	0,05477	84,01%	-15,79%

The R^2 shows how well the predictors explained changes in the response variable. Gradient Boosting is the algorithm with the best R^2. But it's outperformed by Random Forest in terms of Root Mean Squared Error (RMSE). It means that the Gradient Boosting algorithm better explained the variance changes of the "Y" output, but with the presence of larger errors. The negative results in terms of R^2 could be explained by the test set not being representative of the whole dataset. Indeed, the test set is entirely composed of 2022 data, which contains really negative observations due to the really negative events that occurred in the cryptocurrency industry. In this way, the model has not been trained for the outliers that then are present in the test set, and consequently, it struggles to make accurate predictions, resulting in an overall low performance on the test set. If the presence of the outliers is addressed, like in this case, the R^2 could not be an effective way to evaluate the model performance, but it still gives interesting insights into the overall look of the models. As previously explained, looking at scores without having a context is a terrible mistake. One of the most common problems when facing models' scores is the overfitting/underfitting problem. The problem can be identified by looking at the comparison between training and test performances.

Random Forest Regressor: a clear condition of high variance / low bias is readable from the table. The condition is clearly indicated by the negative R^2 and by the wide gap between the train and test R^2. The model fits in a good way the training set but performs badly when exposed to new data (test set). It represents an overfitting situation. A possible solution would be feature reduction, it would force the algorithm to build fewer complex models. Also, increased regularization and hyperparameter tuning can help reduce overfitting.

Gradient Boosting Regressor: it presents the same overfitting situation as the Random Forest Regressor. Furthermore, the test R^2 is even lower in this case. The algorithm presents better performance in training tests than the other two.

MLPRegressor: it represents the worst performer out of the three-algorithm used in this research. The worst performance in the test set is probably due to the highest neural network sensibility and the fact that the test set doesn't well represent the entire dataset.

6.2 *Backtesting*

For backtesting, the Random Forest results have been selected, this choice is due to the lower RMSE and the better R^2 for the test set. In order to check if the variance forecasts done by the Random Forest have a good impact on a strategy formulation, a back-test process has been performed. The forecasted variance values have been used to perform a risk managed momentum strategy. Let $r_{WML_{RF},t}$ be the returns of the random forest risk managed momentum strategy and $\hat{\sigma}_{RF,t}$ the volatility forecast resulted from the random forest implementation. Again, the volatility target used is 12%, the same as the one used in the risk managed momentum strategy performed in step number two.

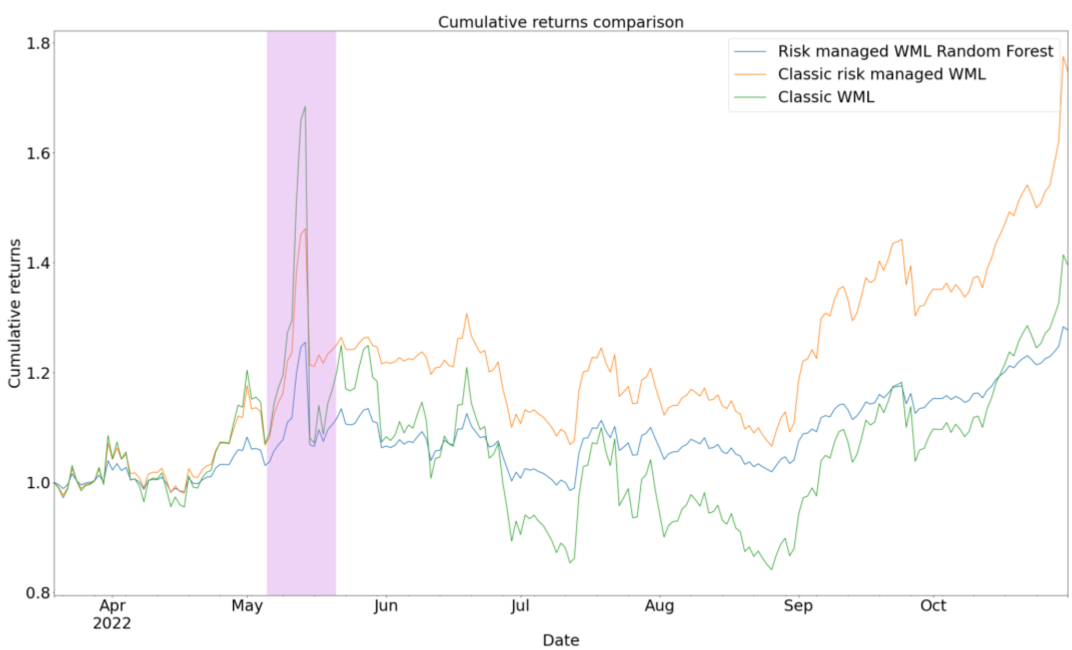
$$r_{WML_{RF},t} = \frac{\sigma_{target}}{\hat{\sigma}_{RF,t}} r_{WML,t}$$

The table is showing the descriptive statistics of the risk managed momentum strategy performed using the random forest variance forecast, the classic risk managed momentum strategy, and a classic vanilla momentum strategy.

Risk managed WML strategies comparison

	Risk managed RandFor	Classic risk managed WML	ClassicWML
Mean (%)	31,73%	71,55%	62,86%
Std. Dev. (%)	27,95%	40,91%	67,02%
Sharpe ratio	1,1351	1,7491	0,9378
Skewness	-2,3305	-0,6104	-2,5472
Kurtosis	23,8980	10,9723	24,0314

The strategy which used the results of the Random Forest algorithm overperformed all the other strategies in terms of lower volatility. The random forest risk managed momentum shows almost half of the volatility of the classic risk managed momentum strategy. Cumulative returns have also been checked to allow a clearer visualization.



= Terra ecosystem crash

The comparison of cumulative returns gives a perfect visualization of what is shown in the previous table. During the Terra ecosystem crash, the random forest risk managed momentum strategy experienced a much smaller crash in comparison to the other two strategies. Overall, we can clearly see how the machine learning implementation for forecasting the volatility of the momentum strategy produces promising results and paves the way for future research.

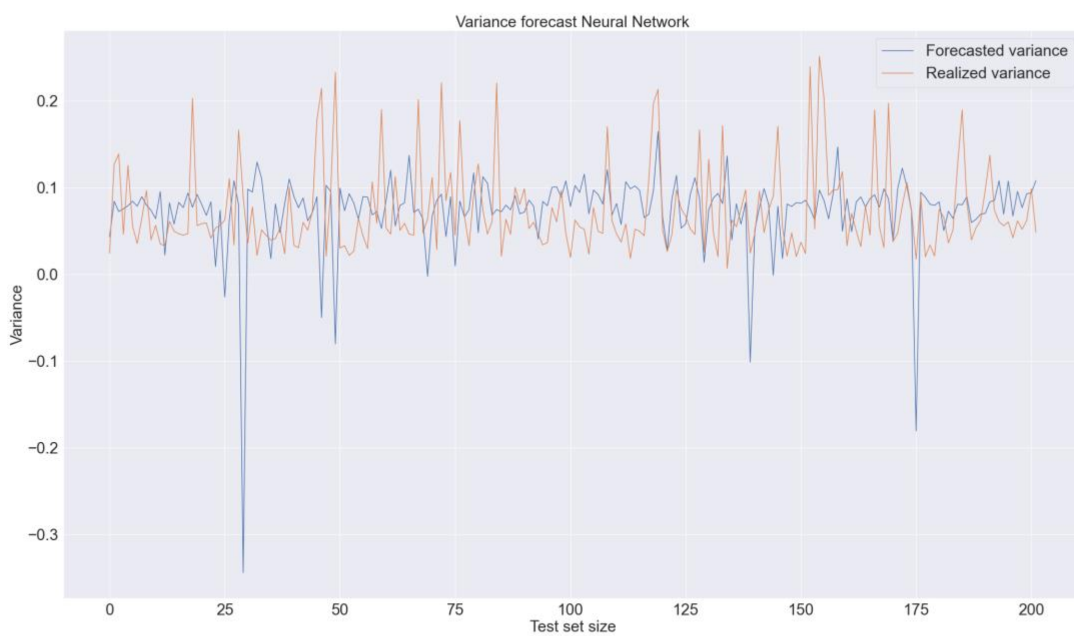
7. Conclusions

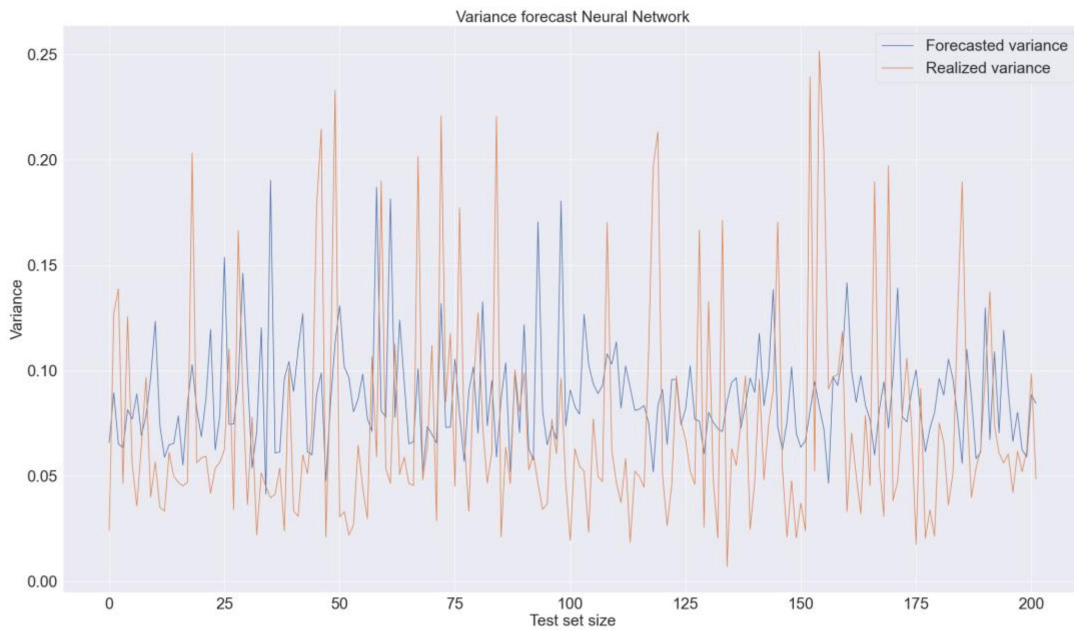
Trend-following techniques are widely used by different types of investors. This study has been developed to check whether these strategies are suitable for cryptocurrency investing at different difficulty levels. The main goal was to test if trend-following strategies could control and reduce cryptocurrencies' volatility and crash sizes and if a machine learning approach could help in the purpose. It has been shown that vanilla trend-following strategies do effectively improve performances by looking at specific indicators. The simple moving average strategy registered the lowest volatility and kurtosis among step 1 (vanilla) strategies. Meanwhile, the classic WML strategy based on the top 125 cryptocurrencies by market capitalization registered the highest Sharpe ratio while maintaining the same volatility as the buy-and-hold strategy. Finally, the machine learning implementation in risk managed momentum strategy shows a good fit for volatility forecasting purposes. In particular, the Random Forest algorithm brought to life a very conservative strategy with almost half of the classic risk managed WML volatility. It also showed a great crash avoidance characteristic as it is shown in the backtesting section above. Overall, it's important to keep in mind that these results come together with some limitations. First, the short dataset timespan brings problems regarding the training and testing processes. As already previously explained, the test set is entirely composed of 2022 data, which represents one of the worst years for cryptocurrencies in terms of performance. Consequently, the already short test set contains a lot of outliers and doesn't represent well the entire dataset. This problem also causes overfitting, because the machine learning algorithms struggle to fit such strange and unbalanced test set. Second, the research has been conducted with raw data and it does not take into consideration transaction costs and borrowing costs resulting from the capital needed for the short positions. Finally, positive correlation among variables has been addressed, this is a common problem with machine learning models. Highly correlated features don’t bring new information to the model, but they increase model complexity, indirectly worsening overfitting.

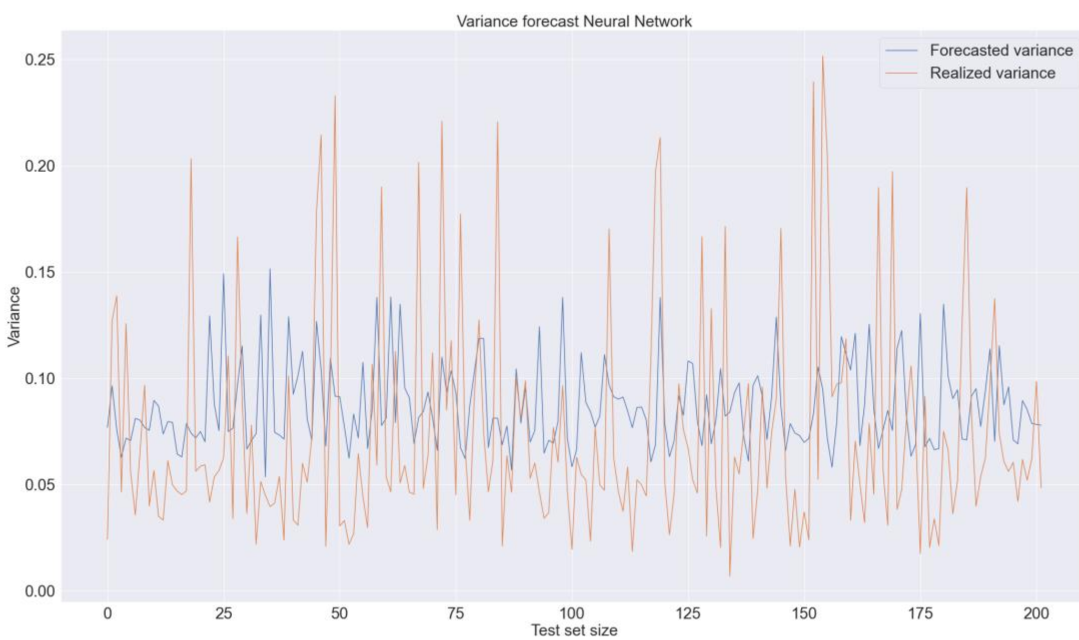
8. Appendix

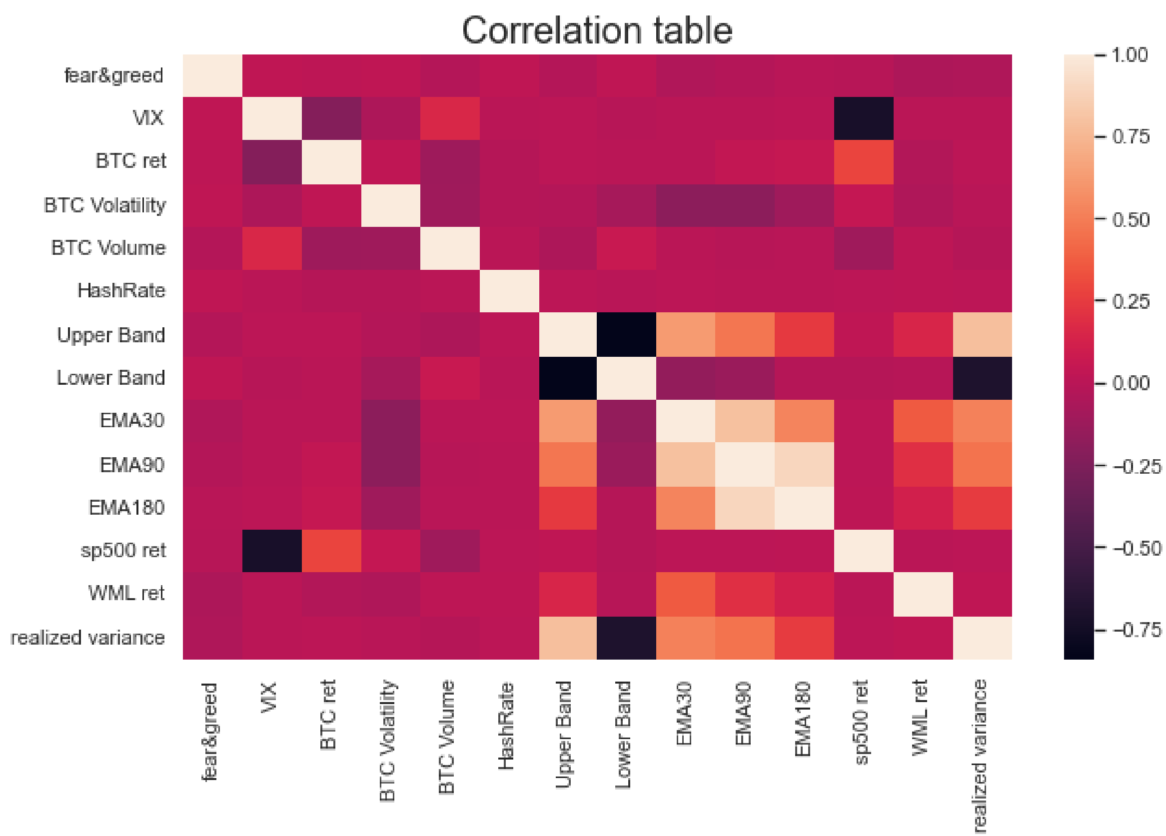
Summary of employed variables

Name of the variable	Group	Frequency	Source
Market capitalizazion top 500 cryptocurrencies	Crypto	Fixed	CoinMarketCap
Prices top 500 cryptocurrencies	Crypto	Daily	YahooFinance
EWMA (30/90/180)	Technical	Daily	Calculated
Fear & Greed Index	Sentiment	Daily	Alternative.me
Bitcoin hash rate	Crypto	Daily	Nasdaq
Bitcoin volumes	Crypto	Daily	Polygon.io
S&P 500 returns	Macro	Daily	FRED
Bollinger bands	Technical	Daily	Calculated
VIX	Macro	Daily	FRED




Variance forecast Neural Network
Forecasted variance
Realized variance
Variance
Test set size
0.25
0.20
0.15
0.10
0.05
0.00
0
25
50
75
100
125
150
175
200


Variance forecast Neural Network
Forecasted variance
Realized variance
Variance
Test set size
0.25
0.20
0.15
0.10
0.05
0.00
0
25
50
75
100
125
150
175
200


Correlation table
fear&greed
VIX
BTC ret
BTC Volatility
BTC Volume
HashRate
Upper Band
Lower Band
EMA30
EMA90
EMA180
sp500 ret
WML ret
realized variance
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

9. Reference List

Ahmed, S., Bu, Z., & Tsvetanov, D. (2018). Best of the best: A comparison of factor models. *Journal of Financial and Quantitative Analysis*, *54*(4), 1713–1758. https://doi.org/10.1017/s0022109018000947

ASNESS, C. L. I. F. F. O. R. D. S., MOSKOWITZ, T. O. B. I. A. S. J., & PEDERSEN, L. A. S. S. E. H. E. J. E. (2013). Value and momentum everywhere. *The Journal of Finance*, *68*(3), 929–985. https://doi.org/10.1111/jofi.12021

Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, *116*(1), 111–120. https://doi.org/10.1016/j.jfineco.2014.11.010

Bhambhwani, S. M., Delikouras, S., & Korniotis, G. M. (2021). Blockchain Characteristics and the Cross-Section of Cryptocurrency Returns. https://doi.org/http://dx.doi.org/10.2139/ssrn.3342842

Bhambhwani, S., Delikouras, S., & Korniotis, G. M. (2019). Do fundamentals drive cryptocurrency prices? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3342842

Biais, B., Bisière, C., Bouvard, M., & Casamatta, C. (2019). The blockchain folk theorem. *The Review of Financial Studies*, *32*(5), 1662–1715. https://doi.org/10.1093/rfs/hhy095

Chen, C. Y. H., HHrdle, W. K., Hou, A. J., & Wang, W. (2018). Pricing cryptocurrency options: The case of CRIX and Bitcoin. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3159130

Christensen, K., Siggaard, M., & Veliyev, B. (2022). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*. https://doi.org/10.1093/jjfinec/nbac020

Cong, L. W., Li, Y., & Wang, N. (2020). Tokenomics: Dynamic adoption and valuation. https://doi.org/10.3386/w27222

Cong, L., Karolyi, G. A., Tang, K., & Zhao, W. (2021). Value Premium, network adoption, and factor pricing of crypto assets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3985631

Dhawan, A., & Putniņš, T. J. (2022). A new wolf in town? pump-and-dump manipulation in cryptocurrency markets. *Review of Finance*. https://doi.org/10.1093/rof/rfac051

FAMA, E. U. G. E. N. E. F., & FRENCH, K. E. N. N. E. T. H. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Hackethal, A., Hanspal, T., Lammer, D. M., & Rink, K. (n.d.). *The Characteristics and Portfolio Behavior of Bitcoin Investors: Evidence from Indirect Cryptocurrency Investments*. Retrieved October 18, 2022, from https://academic.oup.com/rof/article/26/4/855/6478303

Hinzen, F. J., John, K., & Saleh, F. (2022). Bitcoin's limited adoption problem. *Journal of Financial Economics*, *144*(2), 347–369. https://doi.org/10.1016/j.jfineco.2022.01.003

Hong, H., & Stein, J. (n.d.). *A unified theory of underreaction, momentum trading, and overreaction ...* https://onlinelibrary.wiley.com/doi/full/10.1111/0022-1082.00184

Jegadeesh, N., & Titman, S. (2001). Profitability of Momentum Strategies: An evaluation of alternative explanations. *The Journal of Finance*, *56*(2), 699–720. https://doi.org/10.1111/0022-1082.00342

Liu, Y., Tsyvinski, A., & Wu, X. (2019). Common risk factors in cryptocurrency. https://doi.org/10.3386/w25882

Mittnik, S., Robinzonov, N., & Spindler, M. (2015). Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of Banking & Finance*, *58*, 1–14. https://doi.org/10.1016/j.jbankfin.2015.04.003

Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). *Bitcoin and cryptocurrency technologies: A comprehensive introduction*. Princeton University Press.

Pagnotta, E. S. (2021). Decentralizing money: Bitcoin prices and Blockchain Security. *The Review of Financial Studies*, *35*(2), 866–907. https://doi.org/10.1093/rfs/hhaa149

Pagnotta, E., & Buraschi, A. (2018). An equilibrium valuation of bitcoin and decentralized network assets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3142022

Rundo, F., Trenta, F., di Stallo, A. L., & Battiato, S. (2019, December 17). *Machine Learning for Quantitative Finance Applications: A survey*. MDPI. Retrieved November 4, 2022, from https://www.mdpi.com/2076-3417/9/24/5574

Shams, A. (2020). The structure of cryptocurrency returns. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3604322

Sockin, M., & Xiong, W. (2020). A model of cryptocurrencies. https://doi.org/10.3386/w26816

Zhang, W., & Li, Y. (2020). Is idiosyncratic volatility priced in cryptocurrency markets? *Research in International Business and Finance*, *54*, 101252. https://doi.org/10.1016/j.ribaf.2020.101252

Wan, X. (2019). Influence of feature scaling on convergence of gradient iterative algorithm. *Journal of Physics: Conference Series*, *1213*(3), 032021. https://doi.org/10.1088/1742-6596/1213/3/032021

Zhang, W., Li, Y., Xiong, X., & Wang, P. (2021). Downside risk and the cross-section of cryptocurrency returns. *Journal of Banking & Finance*, *133*, 106246.