Portfolio Strategies in Crypto markets: A factor based approach

Submitted By: Mayank Santosh HEDAOO
Supervisor: Co-Pierre GEORG

May 2023

ABSTRACT

In this research, we examine portfolio strategies in the crypto market, specifically focusing on factors such as size, momentum, volatility, trend, and sentiment using various characteristics. Our findings indicate that the size factor portfolio based on market capitalization and the long-short portfolio based on volume outperform other strategies. We also observe that sentiment analysis of crypto news is not effective in accurately predicting market movements, while trend-based strategies demonstrate better performance than sentiment-based approaches. Moreover, we find that a one-factor model fails to adequately explain cross-sectional returns. Overall, this study enhances our understanding of asset pricing dynamics in the coin market and emphasizes the importance of developing theoretical models tailored to the unique characteristics of the crypto ecosystem.

Introduction

In the months from August to October of 2008, the world was falling into a wildfire. The housing market collapsed, Lehman Brothers filed for the largest bankruptcy America had ever seen, Bank of America bought Merrill Lynch, and the U.S. government had to establish a $700 billion Troubled Asset Relief Program (TARP). Amid this turmoil, Bitcoin rose from the ashes like a phoenix, when, on 31st October 2008, Satoshi Nakamoto released the Bitcoin white paper, and Bitcoin.org was registered. This paper served as the genesis for all the blockchain implementations deployed since then. In February 2011, 'Silk Road', an online marketplace was released that provided a rules-free decentralized marketplace for any product one can imagine. This platform used bitcoin as a means of payment. In June 2011, Gawker, an American blog, mentioned "The Silk Road" as "The Underground website where you can buy any drug imaginable" (It is important to know that Bitcoin never endorsed any illegal activities.). This article led to the first spike in Google search for the term 'bitcoin', during which time the bitcoin prices jumped 3 folds in one week, from $10 to $30.

After this spike, different alternatives for bitcoin, termed as altcoins, started to appear, with the likes of Litecoin (LTC) and Namecoin (NMC) being the first, and after Ethereum (ETH) was launched in July 2015, altcoins started to gain popularity. Over the years, the coin market has experienced very rapid growth. CoinMarketCap reports that there are over 22,900 cryptocurrencies with a total market capitalization of over $1 trillion. Liu and Tsyvinski (2021) show that a value-weighted index consisting of all the coins with a market capitalization greater than $1 million, from January 2011 to December 2018, has shown daily, weekly, and monthly average returns of 0.46%, 3.44%, and 20.44% respectively.

There are two prominent views about the coin markets among investors. The first view talks about the coin market as a scam, a bubble, or a Ponzi scheme. It is not backed by any entity, it does not have any central bank, and it is not accepted by tax authorities. Moreover, it is extremely volatile and has a bad reputation. The second view talks about blockchain technology that is underlying the coin market, and how it could disrupt the traditional financial system. Burniske and Tatar (2018) advocate this point of view and ascertain that blockchain is the future of finance while explaining the nuances of blockchain from investment perspective. Thus, if we consider the second case to be true, it becomes necessary to analyze the coin market from the perspective of empirical asset pricing. This will help to set up empirical regularities that can help develop theoretical models and to understand the difference in the theoretical explanations of the asset pricing theories between the coin market and traditional asset classes.

Defining cryptocurrencies and tokens is complex: they can be described as protocols, currency, payment systems, assets, or technology platforms. However, in their current form, cryptocurrencies perform the key economic functions that are performed by a traditional financial system. Aquilina, Frost, and Schrimpf (2023) outline the key functions performed by crypto assets that make them comparable to traditional asset classes (It is important to note that, although stablecoins provide a bridge between the fiat world and the crypto world, crypto assets perform these functions almost exclusively within the decentralized financial ecosystem, and thus, fiat world is insulated from the market failure risk of this ecosystem.): they provide a way of clearing and settling payments, mechanisms for pooling funds, ways to transfer resources through time, across regions and industries, ways to manage uncertainty and control risk, information on prices to facilitate decision making, and ways to deal with incentive problems.

Extensive academic research has been done to study the factors that are important in the valuation of cryptocurrencies. However, like any other asset, the prices of cryptocurrencies and tokens are driven by demand and supply factors.

The supply is usually controlled by the protocol for the currency. For example, the supply of BTC is increasing over time, but at a slower and slower rate, and the total supply is limited to a maximum of 21 million coins. Supply of XPR is decreasing over time because XRP needs to be burnt to fund each transaction. The supply of NT is constant overtime because one pays for transactions using NXT.

The demand for cryptocurrencies and tokens can be driven by different factors. For example, the positive externality of network effect of cryptocurrency adoption (see Pagnotta and Buraschi 2018; Biais, Bisiere, Bouvard, Casamatta, and Menkveld 2018;

Cong, Li, and Wang 2021), the production cost of mining the coins (Cong and He 2019; Sockin and Xiong 2019), traditional asset classes like fiat money (Athey, Parashkevov, Sarukkai, and Xia 2016; Schilling and Uhlig 2019; Jermann 2018), sentiment on news, chatter on social media sites and forums (Stuart Colianni, Stephanie Rosales, and Michael Signorotti 2015), google and Twitter trends (Liu, Tsyvinski 2018), manipulation by large holders using pump and dump schemes (Li, Shin, and Wang 2021) , size factors like market capitalization (Liu, Tsyvinski, and Xiwu 2022), momentum factors like past 1 week returns (Liu, Tsyvinski, and Xiwu 2022), volatility factors (Liu, Tsyvinski, and Xiwu 2022), etc.

In this paper, we first test the mechanisms and predictions of the existing research. Specifically, we test the portfolios formed using size factor using market capitalization, price high and volume as proxy characteristics, momentum factor using last 1 week, 2 weeks, 3 weeks and 4 weeks average weekly returns as proxy characteristics, and volatility factor using price volatility of last 60 days as proxy characteristic. Then we test the effects of meme theory, which suggests that people hold cryptocurrencies because they believe that other people want it too, using sentiment analysis and trend effect on crypto markets.

Size and momentum are among the most studied factors in asset pricing literature. Recent literature on factors affecting cryptocurrencies and tokens provides an explanation for these factors. The size dimension of cryptocurrency can potentially encompass the impact of liquidity. Additionally, the significance of size aligns with the balance between capital gains and the convenience yield as suggested by contemporary theories on cryptocurrencies (Sockin and Xiong in 2018, Prat, Danos, and Marcassa in 2019, and Cong, Li, and Wang in 2021). According to these theories, in a state of equilibrium, larger and more established cryptocurrencies tend to exhibit higher convenience yields, resulting in relatively lower capital gains.

The momentum effect is because of investor overreaction (Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), Sockin and Xiong (2018)). Additionally, contrary to stock and gold where investors are contrarian traders, crypto traders are momentum traders (Kogan, Makarov, Niessner, and Schoar (2022)). One explanation for this type of trading behaviour is that cryptocurrencies and tokens are entirely new investment vehicles whose value is largely dependent on investors' belief about whether there will be a broader market adoption in future (Makarov, Igor and Schoar, Antoinette (2019)). Therefore, investors use price movements as an indicator of future adoption.

Our analysis begins by creating a value-weighted index to evaluate the returns of the cryptocurrency market. We include all coins with a market capitalization exceeding 30 million, covering the period from January 1, 2016, to December 31, 2022. Throughout this time frame, the average daily, weekly, and monthly returns for the coins were 0.20%, 1.41%, and 6.23%, respectively. The corresponding standard deviations for the daily, weekly, and monthly returns were 4.2%, 11.42%, and 25.45%, respectively. The distribution of coin market returns exhibits positive skewness and kurtosis.

In the first part of our thesis, we examine the performance of eight portfolios based on size, momentum, and volatility factors. Each week from January 1, 2016, to December 31, 2022, we categorize cryptocurrencies and tokens into quantile portfolios based on various proxy characteristics, including market capitalization, weekly price high, volume, 1-week momentum, 2-week momentum, 3-week momentum, 4-week momentum, and price volatility over the last 60 days. We then track the subsequent week's performance for each portfolio and calculate the mean returns. Additionally, we create long-short strategies by taking the difference between the 5^{th} and the 1^{st} quantile. Our findings reveal that the long-short strategy returns for Volume and Price Volatility characteristics are statistically significant at confidence levels of 90% and 99%, respectively. Implementing a zero-investment strategy, shorting the highest volume portfolio and longing the lowest volume portfolio generates weekly returns of 3.03% with a volatility of 12.95%. Similarly, a zero-investment strategy, shorting the high volatility portfolio and longing the low volatility portfolio yields weekly returns of 1.19% with a volatility of 12.82%. Moreover, individual quantile portfolios demonstrate statistical significance, with returns changing nearly monotonically. Specifically, for size factor-based portfolios, weekly returns decrease as the quantile increases; for momentum-based portfolios, weekly returns increase with a quantile increase; and for volatility-based portfolios, weekly returns decrease with a quantile increase.

In the second part of our thesis, we utilize Google Trends data for the keywords "bitcoin" and "crypto" along with crypto news sentiment scores to construct weekly and daily portfolios, respectively. The trend portfolio employs the coin market index, while the sentiment portfolio utilizes a value-weighted portfolio of the top 10 coins. Both portfolios exhibit statistical significance at the 99% confidence interval. The sentiment portfolio achieves a mean weekly return of 0.5% with a standard deviation of 4.17%, while the trend portfolio generates a weekly return of 1.18% with a standard deviation of 10.50%.

Finally, we employ the crypto markets CAPM to assess whether a one-factor model with the market as a factor can explain the cross-section of statistically significant portfolios. We also evaluate the performance of all statistically significant portfolios using the Sharpe Ratio, and alpha that represents the difference between expected return over the risk-free rate and realized return over risk-free rate. Our findings indicate that the one-factor model performs poorly and fails to explain the cross-sectional returns for any of the portfolio strategies. The R-squared values range from 0.008 for the long-short volume-based portfolio to 0.301 for the sentiment-based portfolio. These results align with those observed in traditional asset classes (Liu, Tsyvinski, and Wu, 2022). Additionally, we find that the market capitalization-based size factor portfolio and the volume-based long-short size factor portfolio exhibit maximum Sharpe ratios of over 19%. Conversely, price volatility based long-short portfolio and sentiment-based portfolios display minimum Sharpe ratios of 3.69% and 1.59% respectively.

Literature Review

Some of the earliest literature on bitcoin and cryptocurrencies believed that bitcoin largely fails to satisfy the functions of a currency, i.e., medium of exchange, store of value and unit of account. It has higher volatility than any other currency in the world, and there are no means to hedge the risk associated with bitcoin (Yermack (2015)). However, things changed after Ethereum was launched in July 2015, and DApps and smart contracts were introduced in the crypto world. After this, many researchers changed their view about cryptocurrencies and saw blockchain as a force that could disrupt the financial system (Burniske, Tatar (2018)).

Several articles and papers studied the individual behavioural patterns of cryptocurrencies and tokens. Stoffels (2017) conducted research on the momentum of cryptocurrencies using a cross-sectional approach. Hu, Parlour, and Rajan (2018) demonstrated a correlation between individual cryptocurrency returns and Bitcoin returns. Borri (2019) discovered that individual cryptocurrencies are susceptible to tail risks in the cryptocurrency market. Griffin and Shams (2020) focused on studying the manipulation of Bitcoin prices, while Corbet et al. (2019) investigated cryptocurrencies as financial assets. Sangyup Choi and Junheok Shin (2022) studied the inflation hedging properties of bitcoin and discovered that bitcoin is not a safe haven asset, but its prices do not decrease against policy uncertainty shocks, and they increase against increasing inflation expectations. Glaser et al. (2014) studied the user intentions behind changing their domestic currency to digital currency and discovered that users, especially uninformed users perceive bitcoin as an alternate investment vehicle rather than an alternate transaction system.

However, the issue of determining the value of cryptocurrencies and tokens persists. In the case of tokens representing ownership of a tangible asset like 1 ounce of gold, the token's price is expected to closely follow the price of the underlying asset. Nevertheless, cryptocurrencies lack such backing or claim to any specific entity. Consequently, similar to other assets, we can infer that the prices of cryptocurrencies and tokens are influenced by the forces of supply and demand.

Emissions, inflation/deflation, and distributions are the main drivers of supply (Nat Eliason (2021)). A token will increase its value if fewer of them exist and decrease if more of them exist. For example, BTC was created with a supply of 21 million, the last of which will be mined in the year 2140. Roughly, 19 million already exist and only 2 million are about to be mined in the next 120 years. Therefore, we may not expect any serious inflationary pressure to bring down the value of BTC. On the other hand, ETH has no limit on the supply. Currently, over 118 million ether coins are circulating. But, after recent adjustments to net emission, ETH will reach a stable supply of around 120 million.

Although the supply of any cryptocurrency is tied to an algorithm, the ‘fixed’ nature of supply is debatable. For instance, in the case of BTC, if 51% of the network decides to switch its algorithm to adopt the change, then a new fork is created which has an increased supply.

Distributions or allocation of coins are different for different protocols. So, to know more about the allocations of a coin, investors must ask how many investors hold most coins and when will they be unlocked. And did the protocol give most of its coins to the community? For example, the concentration of BTC is highly concentrated. About 2.17% of the accounts hold 93.2% of BTC.

The demand for cryptocurrencies can be explained by ROI, Meme theory and Game theory (Nat Eliason (2021)). Return on investment or ROI in this case is the amount of income or cash flows the coin is able to generate by simply holding the coin. For example, investors can stake ETH to get more ETH at an ROI of about 5%. ROI is important to consider, because if the coin does not have any intrinsic cashflows, then it is harder to justify holding it.

Meme theory suggests that another reason why people might want to hold cryptocurrencies is simply the belief that other people want it too. One way to evaluate this theory is to get a feel of what is happening on community platforms like Discord or Twitter. Belief in future value is one of the most powerful drivers of demand. Colianni et. al (2023) find that sentiment analysis of day-to-day tweets on bitcoin can predict the bitcoin movements with 95% accuracy.

The game theory considers what additional elements in tokenomics might help generate demand for the token. A good example of game theory is the lockup protocol. The lockup protocol creates an incentive for locking tokens in a smart contract to generate greater rewards.

Many recent papers and articles have also focused on the risk and return factors in crypto markets and various trading strategies. Yukun Liu (2020) tests if the returns on cryptocurrencies show signs of network effect and production effect. To measure the network effect, they use the number of wallet users, number of active addresses, number of transaction counts and number of payment counts. They find that coin market returns are positively and significantly exposed to the network effect. The economic explanation they give to justify the results is that cryptocurrency prices reflect current cryptocurrency adoption and the expected future network growth.

On the other hand, they study the production effect to understand the implications of miners' production problem on cryptocurrency prices, since the production of cryptocurrencies requires electricity and computing power. They find that cryptocurrency returns are not significantly related to the production factor, just like the cost of mining gold has nothing to do with the price of gold.

Makarov and Schoar(2020) study arbitrage trading and price formation in crypto markets. They find that crypto markets lack any provisions to ensure that investors receive the best price, which increases the role of arbitrageurs who trade across different markets. They show that there are large and recurring deviations in bitcoin prices across exchanges that often persist for several hours. The price deviations are much larger across countries than within the same country. The daily amount of potential arbitrage profits in the period from Dec 2017 to Feb 2018 was often more than $75 million.

Li, Shin and Wang (2022) study the trading strategies that are banned in traditional markets like manipulation theories and pump & dump (P&D) schemes in crypto markets. They find that manipulators often organize “pump” groups using encrypted social media platforms like Telegram. Typical P&Ds last only for several minutes as opposed to stock markets, where they last for several months, and identifying them is much easier in crypto markets than in stock markets since, in the crypto markets, manipulators organize pump groups, advertise on social media, and disclose their pump plans in real-time.

Yukun Liu and Xi Wu (2022) analyzed 24 characteristics in 4 groups of risk factors: size, momentum, volume and volatility. They found that returns of 10 characteristics – market capitalization, price, maximum price, returns of past 1 week, 2 weeks, 3 weeks, 4 weeks & 1 to 4 weeks, price volume and standard deviation of price volume – significantly explain the returns in crypto markets. A parsimonious 3-factor model of market factor, size factor and momentum factors formed using Fama and French (1993) methodology can explain the cross-sectional returns of all the 10 significant portfolios. Premium associated with size factor is because the size factor captures the illiquidity premium of the market, whereas premium associated with momentum factor is because of the investor overreaction/ underreaction to the initial rally. Kogan, Makarov, Niessner and Schoar (2022) find that investors are momentum traders in crypto assets and contrarian in stocks and golds. Cryptocurrencies are an entirely new investment vehicle whose future value largely depends on investors' beliefs about the future of crypto assets. Therefore, the investors might use momentum as an indicator of future adoption.

Our goal in this thesis is to test the portfolios formed using size, momentum and volatility factors. Further, we extend the work of Colianni et al. (2023) and use crypto news and Google trends to form and test trading strategies.

Data and Basic Characteristics

Cryptocurrencies and tokens trade on more than one exchange, and prices and liquidity across exchanges may defer (Recall Makarov and Schoar (2020)). We gather trading data for every cryptocurrency listed on Coinmarketcap.com, a prominent platform renowned for its comprehensive cryptocurrency price and volume information. Coinmarketcap.com consolidates data from more than 200 prominent exchanges, offering daily insights into the opening, closing, high, and low prices, as well as the volume and market capitalization (in USD) of the majority of cryptocurrencies. To determine the price of each cryptocurrency featured on the website, a volume-weighted average is calculated based on the prices reported across various markets. To qualify for listing, a cryptocurrency must fulfill specific criteria, including being traded on a public exchange with an accessible application programming interface (API) that provides data on the last traded price and the trading volume over the past 24 hours. Additionally, the cryptocurrency must have nonzero trading volume on at least one supported exchange to enable the calculation of a meaningful price.

We use daily close price to calculate weekly returns for all the coins having market capitalization of more than $30 million from January 2016 to December 2022. Additionally, we require coins to have information on daily open, close, high, low, volume and market capitalization. Table 1 represents the summary statistics of the data. The number of coins that satisfy our filters increase from 12 in 2016 to 217 in 2022. The mean of market capitalization and volume for all the years is substantially greater than the median, meaning that the distribution is heavily skewed.

Table 1
Summary Statistics

This table reports the number of coins, the mean and median of market capitalization, and the mean and median of daily price volume for all the years. Rows represent the year, and columns represent the no. of coin, and, mean and median of volume and market capitalization.

		Volume		**Market Capitalization**	
	No. of Coins	Mean(millions)	Median(millions)	Mean(millions)	Median(millions)
2016	12	9.43	0.39	860.7	37.67
2017	21	224.64	16.07	5382.18	394.23
2018	51	321.73	10.84	5745.05	503.17
2019	81	679.34	8.02	2669.63	130.91
2020	115	1071.25	13.74	2941.06	127.51
2021	176	1467.82	74.95	11274.24	956.24
2022	217	652.92	22.12	6320.18	345.94

Further, we construct a value weighted index for all the underlying coins using daily close price and calculate weekly returns for the index. Return characteristics of coin market index, BTC, ETH and XRP are presented in Table 2. Coin market mean weekly returns (1.41%) are similar to BTC (1.52%), but lower than ETH (3.29%) and XRP (2.87%). However, the volatility of ETH (17.94%) and XRP (22.23%) is much higher than that of coin market (11.42%) and BTC (10.33%). Sharpe ratio of ETH (18.23%) is the highest followed by BTC (14.52%). All of them have positive skewness and kurtosis.

Table 2
Return Characteristics

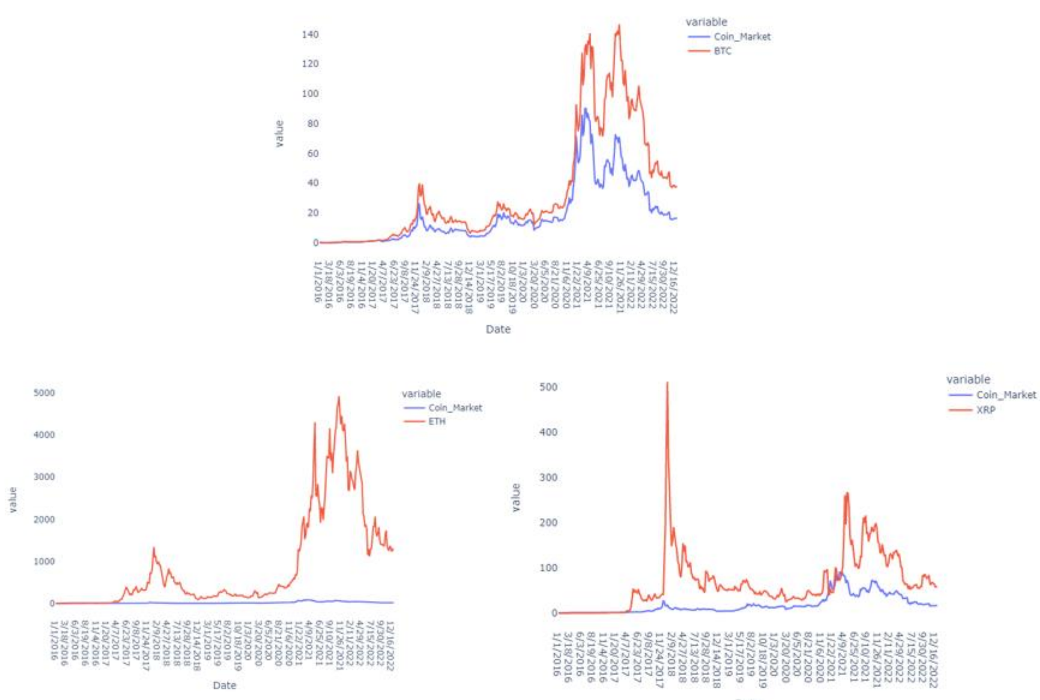
This table reports the characteristics of coin market, BTC, ETH and XRP weekly returns from 1st January 2016 to 31st December 2022. One month Treasury rate has been considered as the risk-free rate to calculate Sharpe ratio.

	Mean	Median	Std	Skew	Kurt	Sharpe Ratio
Coin Market Returns	1.41%	0.74%	11.42%	0.69	3.72	12.17%
BTC	1.52%	1.09%	10.33%	0.31	2.42	14.52%
ETH	3.29%	0.55%	17.94%	2.51	14.86	18.23%
XRP	2.87%	-1.41%	22.23%	3.72	22.59	12.82%

We now document the time series properties of coin market returns. Figure 1 plots the price movements of coin market and BTC, ETH and XRP. The cumulative returns of BTC and coin market follow the same path and are highly correlated. However, the cumulative returns of XRP and ETH are much higher and volatile than the coin market.

Figure 1
Coin market returns and BTC, ETH, XRP

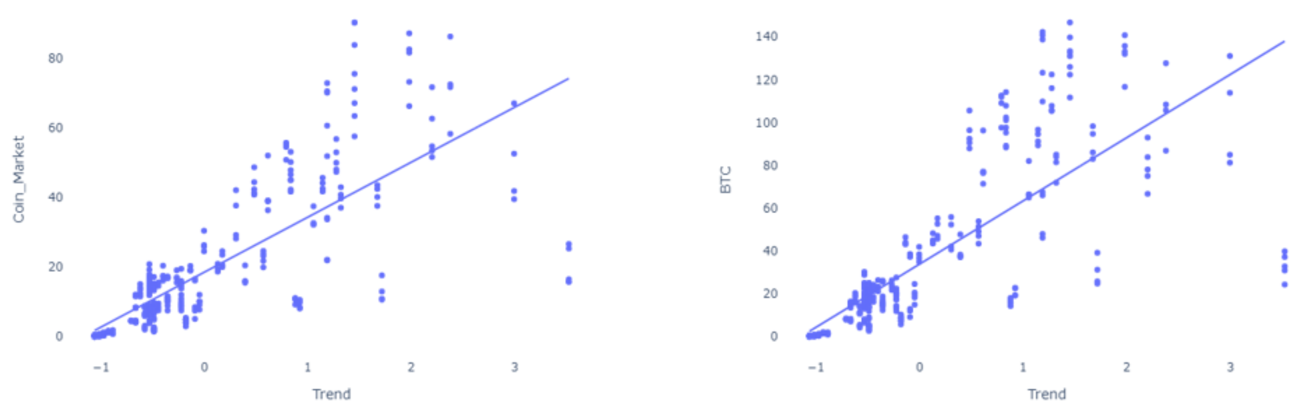
The figure plots journey of $1 invested in coin market, BTC, ETH and XRP.



For analyzing the trend factor, we download the google search data series for keywords "bitcoin" and "crypto currency" from trends.google.com. Google Trends is a powerful tool that provides valuable insights into the popularity and interest levels of specific searched terms over time. By analyzing the billions of searches conducted on Google, it helps users understand and visualize the relative popularity of various topics, keywords, or queries. Google Trends allows users to explore the volume of search queries by region, language, and time - period, giving them the ability to track the rise and fall of interest in specific subjects. Figure 2 plots the scatter plot between trend and cumulative returns for coin market and BTC. There exists a noteworthy correlation between the Google trends cumulative returns of coin market and BTC. Moreover, trends have the potential to elucidate the movement of cumulative returns for both the coin market and BTC.

Figure 2
Scatter plot

The figure plots scatter plot between trend (scaled to have mean of 0 and volatility of 1) and coin market and BTC, along with OLS trendline.



In addition, we use crypto news articles' headlines and summaries sourced from cointelegraph.com to assess the sentiment factor. Cointelegraph.com is a highly regarded and objective digital media platform that specializes in delivering comprehensive coverage of blockchain technology, cryptocurrency assets, and emerging trends in the fintech industry. As a leading independent source, Cointelegraph offers a wide range of news articles, analysis, and insights on the latest developments within the blockchain and crypto space. With a commitment to provide unbiased information, Cointelegraph serves as a valuable resource for investors, enthusiasts, and professionals seeking reliable and up-to-date information about cryptocurrencies, decentralized technologies, and the broader impact of blockchain on various industries. Its extensive archive of news articles makes it a valuable reference for researchers and analysts in understanding the evolution and trends of the crypto landscape. We have gathered a total of 7,172 news articles from January 1, 2016, to December 31, 2022. Each day, we compile all the news summaries into a single collection and conduct sentiment analysis on them.

Cryptocurrency Factors & Characteristics

In the realm of cryptocurrencies and tokens, existing academic literature has proposed various cryptocurrency-specific factors that can influence cryptocurrency prices and forecast their returns. In this section, we will delve into the implications of these factors. We commence by examining the size, momentum, and volatility factors put together by Liu, Tsyvinski, and Wu (2022) to elucidate the variation in crypto returns across different assets. Subsequently, we evaluate the trend factor by utilizing Google Trends data and conducting sentiment analysis on crypto news obtained from Cointelegraph.com. Next, we employ the coin market Capital Asset Pricing Model (CAPM) on all portfolios to identify which portfolio returns can be significantly projected using a one-factor model based on coin market excess returns. Finally, we compare all the portfolio strategies that exhibit significant predictability by the coin market using metrics such as Sharpe ratio, alpha, mean, and volatility.

I. Size

We use market capitalization, volume, and weekly high as characteristics to analyze the size factor. Each week, we sort the individual cryptocurrencies and tokens into quantile portfolios based on the value of the characteristic. Additionally, we also form long – short zero investment portfolios by subtracting 1st quantile from 5th quantile. We track the returns of each portfolio in the next week. Further, we test whether the excess portfolio returns over risk-free rate can be predicted by excess coin market returns over risk-free rate, i.e., we apply coin market CAPM to all the portfolios. Table 3 reports the results from our analysis.

Table 3
Size Strategy Returns

This table reports the predictors based on size-based portfolios and their t-statistics when coin market one factor model, i.e., coin market CAPM is applied. The pricing model applied is (CMKT is coin market returns):

$$Ri - Rf = \alpha + \beta i\ (CMKT - Rf)$$

	Quintiles					
	1	2	3	4	5	5 - 1
Market Cap based portfolios						
Beta	0.5455	0.5617	0.5910	0.6277	0.5842	0.0388
t - stat	8.160	9.111	10.383	12.689	12.552	0.797
Volume based portfolios						
Beta	0.5070	0.5152	0.6537	0.5885	0.5996	0.0927
t-stat	7.417	8.169	11.340	11.243	12.499	1.678
Weekly High Price						
Beta	0.6469	0.5899	0.5089	0.5071	0.6072	-0.0396
t-stat	8.956	10.188	9.657	8.936	12.443	-0.713

We find that market returns can significantly predict all the quintile portfolios at 99% confidence level, and long-short portfolio based on volume characteristic at 90% confidence level. However, long-short zero investment portfolios based on market capitalization and weekly high characteristics cannot be predicted by market returns.

II. Momentum

We use the last 1-week, 2-week, 3-week, and 4-week weekly returns as characteristics to analyze the momentum factor. Each week, we sort the individual cryptocurrencies and tokens into quantile portfolios based on the value of the characteristic and rebalance them every week. Additionally, we also form long – short zero investment portfolios by subtracting 1st quantile from 5th quantile. We track the returns of each portfolio in the next week. Then, we test whether the excess portfolio returns over risk-free rate can be predicted by excess coin market returns over risk-free rate, i.e., we apply coin market CAPM to all the portfolios. Table 4 reports the results from our analysis. We find that market returns can significantly predict all the quintile portfolios at 99% confidence level. However, long-short zero investment portfolios cannot be predicted by market returns.

Table 4
Momentum Strategy Returns

This table reports the predictors based on momentum-based portfolios and their t-statistics when coin market one factor model, i.e., coin market CAPM is applied. The pricing model applied is (CMKT is coin market returns):

$$Ri - Rf = \alpha + \beta i\ (CMKT - Rf)$$

	Quintiles					
	1	2	3	4	5	5 - 1
1 week momentum						
Beta	0.5367	0.5576	0.6349	0.5556	0.5833	0.0467
t-stat	9.425	10.716	10.763	9.805	8.952	0.805
2-week momentum						
Beta	0.5999	0.5652	0.6056	0.5557	0.5666	-0.0331
t-stat	10.385	10.984	10.557	10.188	8.522	-0.624
3-week momentum						
Beta	0.5336	0.5987	0.5914	0.5315	0.5920	0.0585
t-stat	9.644	11.344	10.797	9.701	9.007	1.085
4-week momentum						
Beta	0.5514	0.5918	0.5561	0.5492	0.6027	0.0514
t-stat	9.632	11.694	10.147	9.308	9.444	0.964

III. Volatility

We use the last 60 days price volatility as a characteristic to analyze volatility factor of crypto returns. Each week, we sort the individual cryptocurrencies and tokens into quantile portfolios based on the value of the characteristic and rebalance them every week. Additionally, we also form long – short zero investment portfolios by subtracting 1st quantile from 5th quantile. We track the returns of each portfolio in the next week. Then, we test whether the excess portfolio returns over risk-free rate can be predicted by excess coin market returns over risk-free rate, i.e., we apply coin market CAPM to all the portfolios. Table 5 reports the results from our analysis. We find that market returns can significantly predict all the portfolios at 99% confidence level.

Table 4
Volatility Strategy Returns

This table reports the predictors based on volatility-based portfolios and their t-statistics when coin market one factor model, i.e., coin market CAPM is applied. The pricing model applied is (CMKT is coin market returns):

$$Ri - Rf = \alpha + \beta i \, (CMKT - Rf)$$

	Quintiles					
	1	2	3	4	5	5 - 1
Price-volatility						
Beta	0.6528	0.4322	0.3714	0.3555	0.274	0.1224
t-stat	12.023	11.108	10.201	10.568	8.238	2.503

IV. Trend

The theoretical literature on cryptocurrencies has indicated a potential connection between investor attention and future cryptocurrency returns, as demonstrated by studies such as Sockin and Xiong (2019). In this section, we examine the role of investor attention in forecasting cryptocurrency returns using weekly search volume results from Google trends. To form portfolio using trend, each week we calculate the difference between the search volume from current week and last week. If the difference is positive, then we go long on the coin market, if the difference is zero, we skip that week, and if the difference is negative, we go short. We track the returns from the portfolio the following week. Then, we test whether the excess portfolio returns over risk-free rate can be predicted by excess coin market returns over risk-free rate, i.e., we apply coin market CAPM to the portfolio. The model that we apply is:

$$Ri - Rf = \alpha + \beta i \, (CMKT - Rf)$$

We find that market returns can significantly predict the trend portfolio at 99% confidence level (Beta = 0.3838, t-stat (Beta) = 3.478).

V. Sentiment

According to Shiller (2019), in order to enhance our comprehension of the economy and financial markets, economists should expand their focus beyond conventional economic indicators and include narratives that impact both individual and collective economic behavior. While market participants are affected by what they observe, hear, and discuss, narratives are abstract and challenging to quantify. In this section, we employ textual analysis techniques, specifically Natural Language Processing (NLP) methods, on a collection of news articles obtained from cointelegraph.com.

The selected timeframe for the articles spans from January 1, 2016, to December 31, 2022, and contains a total of 7,172 news articles. To facilitate analysis, we aggregate the headlines and summaries from each day's articles into a unified collection, creating reservoirs of news articles. Subsequently, we employ the following data preprocessing techniques to clean the text for analysis:

1. Sentence Tokenization: Using the nltk library, we divide the text for each day into individual sentences.
2. Normalizing: In this step, we eliminate numbers and non-words from the text.
3. Stop Words: Certain words, such as ‘is’, ‘a’, ‘the’, carry minimal information and can impede analysis speed. Hence, we eliminate them from the text using stop words provided by nltk library.
4. Lemmatization: Many words are variations of their root forms. For instance, ‘running’, ‘ran’ and ‘run’ all convey the same meaning. In this step, we convert each word to its respective root form.

Further, we opted to utilize dictionary-based methods and initiated our analysis with the VaderSentiment algorithm to calculate sentiment scores for our collection of news articles. This particular approach employs a pre-defined dictionary and a set of rules to perform sentiment analysis, with a specific focus on measuring sentiments in social media content. The algorithm provides a sentiment score ranging from -1 to +1, where a score of -1 indicates highly negative sentiment, and a score of +1 signifies highly positive sentiment.

We utilize the sentiment scores obtained as inputs for two types of models: a classification model designed to predict coin market movement (+1 if market goes up and -1 if market goes down), and a trading model intended to generate profitable signals.

For our first model, we make use of supervised learning algorithms like Logistic Regression and Random Forest to test whether the sentiment scores can predict market movements.

1. Logistic regression is a statistical technique that is commonly used for binary classification tasks where the target variable only has two possible outcomes. For instance, it can be applied to detect cancer. Logistic regression models the relationship between the independent variables and the probability of the event occurrence. The probability is calculated by applying sigmoid function to the linear regression output.
2. The Random Forest model is an ensemble of tree models, specifically a combination of multiple machine-learning models, making it a powerful successor to tree-based models. It has been demonstrated by Breiman (1999) that Random Forest Models offer ease of implementation, accurate predictions, and most significantly, the ability to handle a large number of input variables without overfitting.

We test the fitness of our selected machine learning algorithms using metrics like accuracy, confusion matrix and AUC-ROC scores. Accuracy in the number of correct predictions over total number of predictions:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}$$

Where TN = True Negatives, TP = True Positives, FN = False Negatives, FP = False Positives. AUC is the area under the curve, and the larger the value of AUC, the more effective the classification algorithm. AUC is also written as AUC-ROC (Area Under the Receiver Operating Characteristics). The value for AUC ranges from 0 to 1, and a model whose predictions are wrong always has an AUC of 0.0, and one whose predictions are correct always has an AUC of 1. It thus measures the quality of predictions.

Table 5 reports the accuracy, AUC-ROC score, and confusion matrix for both the learning algorithms. We find that sentiment scores calculated using VaderSentiment algorithm cannot predict the market movements accurately, and can achieve an accuracy of approximately 50%.

Table 5
Detailed Metrics

This table reports the accuracy, AUC-ROC scores and confusion metrics for Random Forest and Logistic Regression machine learning algorithms ran on sentiment scores to predict market movements.

		Random Forest		**Logistic Regression**	
Accuracy		0.52		0.53	
AUC-ROC Score		0.51		0.54	
Confusion Matrix		Negative (Predicted)	Positive (Predicted)	Negative (Predicted)	Positive (Predicted)
	Negative (Actual)	88	124	98	114
	Positive (Actual)	84	134	86	132

In our second model, we utilize sentiment scores to construct a portfolio strategy. We observe that the news articles solely pertain to the top 10 coins, namely 'BTC', 'ETH', 'BNB', 'XRP', 'ADA', 'DOGE', 'MATIC', 'SOL', 'DOT', and 'LTC'. Therefore, we create a value-based index comprising these top 10 coins, which will serve as the foundation for our portfolio strategy. The daily average return of this index is 0.2%, with a daily volatility of 4.08%. To establish the portfolio based on sentiment scores, we adopt the following approach:

1. If the sentiment score exceeds 0.5, we take a long position in the top-10 coin market index.
2. If the sentiment score falls below -0.5, we take a short position.
3. All other values of the sentiment scores are disregarded in this strategy.

Since news is published daily, we track the performance of our portfolio daily. Then, we test whether the excess portfolio returns over risk-free rate can be predicted by excess coin market returns over risk-free rate, i.e., we apply coin market CAPM to the portfolio. The model that we apply is:

$$Ri - Rf = \alpha + \beta i \, (CMKT - Rf)$$

We find that market returns can significantly predict the trend portfolio at 99% confidence level (Beta = 0.4646, t-stat (Beta) = 20.295).

Portfolio Analysis

In this section, we analyze the performance of various portfolios whose returns can be significantly predicted by market returns. These portfolios encompass different factors such as size (market capitalization, price high, volume), momentum (1-week momentum, 2-week momentum, 3-week momentum, 4-week momentum), volatility (price volatility), volume-based long-short, price volatility-based long-short, as well as trend and sentiment-based portfolios. However, due to the portfolios' monotonic increase or decrease, specifically, we choose the 1^{st} quantile portfolio for size factor and volatility factor-based portfolios, and the 5^{th} quantile portfolio for momentum factor for comparison.

In order to assess various portfolios, we employ several metrics. First, we utilize alpha derived from the coin market CAPM. This involves calculating the difference between the expected return over the risk-free rate and the realized return over the risk-free rate. Additionally, we employ the t-statistics of the alpha to determine its significance. To evaluate the extent to which the cross-sectional returns of the portfolio can be explained by the coin market returns, we consider the R-squared value from the CAPM regression. Furthermore, we compare the risk-adjusted returns of the portfolio using the Sharpe ratio. To calculate Sharpe ratio, we utilize the 1-month U.S. Treasury Bill rate as risk-free rate. Table 6 reports the values for the metrics under consideration for all the portfolios, and Figure 3 provides the visual representation of the cumulative returns of the portfolios.

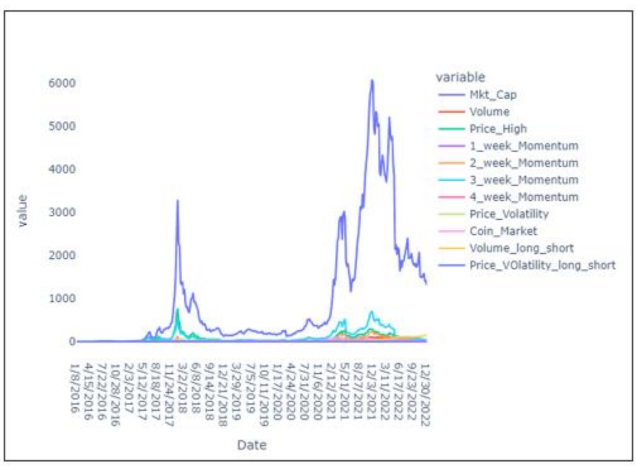
Table 6
Summary Statistics

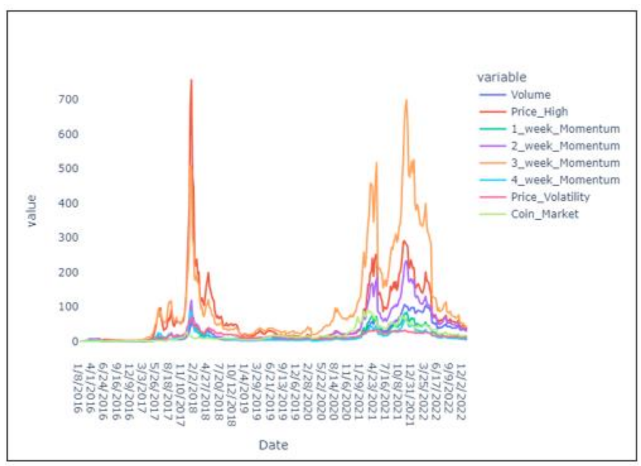
The table presents the alpha obtained from CAPM, the R-squared value from CAPM, the average weekly returns, weekly volatility, and Sharpe ratio for all the portfolios that can be forecasted using coin market returns.

Portfolio Strategies	Alpha	t(alpha)	R square	Mean Weekly Returns	Volatility	Sharpe Ratio
Market Cap	0.0269	3.587	0.1	3.12%	15.50%	19.99%
Volume	0.0196	2.561	0.1	2.36%	15.64%	15.00%
Price High	0.0186	2.304	0.1	2.37%	17.02%	13.84%
1-week momentum	0.0155	2.121	0.1	2.01%	13.54%	12.97%
2-week momentum	0.0183	2.461	0.1	2.28%	14.03%	14.57%
3-week momentum	0.0193	2.616	0.1	2.39%	13.22%	15.33%
4-week momentum	0.0132	1.85	0.1	1.80%	13.67%	11.72%
Price Volatility	0.0069	1.686	0.2	1.03%	9.15%	11.11%
Volume Long-Short	0.0238	3.836	0.008	3.03%	12.95%	19.12%
Price Volatility Long-Short	0.0059	0.947	0.017	1.19%	12.82%	3.69%
Trend	0.003	0.256	0.142	1.18%	10.50%	11.06%
Sentiment	0.0011	0.981	0.301	0.50%	4.17%	1.59%

Figure 3
Cumulative Portfolio Returns

The left figure shows the journey of $1 invested in all the portfolios except trend and sentiment-based portfolios. The market capitalization-based portfolio has clearly shown higher cumulative returns than any other portfolio. The right figure shows cumulative returns of all portfolios except market capitalization, trend and sentiment portfolios.





After conducting our analysis, we have discovered some noteworthy findings regarding the performance of different portfolios predicted by coin market returns. Market capitalization stands out as the top-performing portfolio, showcasing exceptional results with a Sharpe ratio of 19.99% and an alpha of 0.269. Following closely behind is the volume long-short zero investment portfolio, which exhibits a Sharpe ratio of 19.12% and an alpha of 0.0238.

In terms of risk, the portfolios based on weekly high prices and volume present higher levels of volatility, with standard deviations of 17.02% and 15.64% respectively. These portfolios entail greater fluctuation in returns compared to others in the analysis.

Additionally, we find that none of the portfolios' cross-sectional returns can be explained by the one-factor model of the coin market, namely the coin market CAPM. This outcome is consistent with the findings observed for other assets, suggesting that the selected portfolios do not exhibit a strong relationship with the single-factor model.

Overall, the results highlight the dominant performance of the market capitalization based portfolio, the relatively high-risk nature of weekly high price based and volume based portfolios, and the limited explanatory power of the coin market CAPM across all analyzed portfolios.

Bibliography

1. Liu, Yukun, and Aleh Tsyvinski, 2021, Risks and Returns of Cryptocurrency, . Ed. Itay Goldstein The Review of Financial Studies 34, 2689–2727.
2. Burniske, Chris, and Jack Tatar, 2018, Cryptoassets: The innovative investor's guide to Bitcoin and beyond, .
3. Aquilina, Matteo, Jon Frost, and Andreas Schrimpf, 2023, Decentralised Finance (DeFi): A Functional Approach, SSRN Electronic Journal.
4. Pagnotta, Emiliano, and Andrea Buraschi, 2018, An Equilibrium Valuation of Bitcoin and Decentralized Network Assets, SSRN Electronic Journal.
5. Biais, Bruno, Equilibrium Bitcoin Pricing, .
6. Cong, L. W., Y. Li, and N. Wang. 2021. Tokenomics: Dynamic adoption and
7. valuation. Review of Financial Studies 34 (3): 1105{1155.
8. Cong, L. W., and Z. He. 2019. Blockchain disruption and smart contracts. Review of
9. Financial Studies 32:1754{1797.
10. Sockin, M., and W. Xiong. 2019. A model of cryptocurrencies. Working Paper,
11. Princeton University.
12. Athey, S., I. Parashkevov, V. Sarukkai, and J. Xia. 2016. Bitcoin pricing, adoption,
13. and usage: theory and evidence. Working Paper, Stanford University.
14. Schilling, L., and H. Uhlig. 2019. Some simple bitcoin economics. Journal of Monetary
15. Economics 106:16{26.
16. Jermann, U. J. 2018. Bitcoin and cagan's model of hyperination. Available at SSRN
17. 3132050.
18. Colianni, Stuart, Stephanie Rosales, and Michael Signorotti, Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis, .
19. Li, Tao, Donghwa Shin, and Baolian Wang, 2018, Cryptocurrency Pump-and-Dump Schemes, SSRN Electronic Journal.

20. LIU, Y., TSYVINSKI, A. and WU, X. (2022), Common Risk Factors in Cryptocurrency. The Journal of Finance, 77: 1133-1177. https://doi.org/10.1111/jofi.13119
21. Sockin, M, and W Xiong, 2018, A model of cryptocurrencies. Princeton University, Unpublished manuscript.[Google Scholar].
22. Prat, Julien, Vincent Danos, and Stefania Marcassa, 2021, Fundamental pricing of utility tokens, .
23. Kogan, S., I. Makarov, M. Niessner, and A. Schoar. 2022. Are cryptos di_erent?
24. evidence from retail trading. Available at SSRN 4289513.
25. Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, Journal of financial economics 49, 307–343.
26. Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, the Journal of Finance 53, 1839–1885.
27. Sockin, Michael, and Wei Xiong, 2023, A model of cryptocurrencies, Management Science.
28. Makarov, Igor, and Antoinette Schoar, 2019, Price discovery in cryptocurrency markets, .
29. Yermack, David, 2015, Is Bitcoin a real currency? An economic appraisal, Handbook of digital currency (Elsevier).
30. Stoffels, J. 2017. Asset pricing of cryptocurrencies and momentum based patterns. Working Paper, Erasmus
31. School of Economics.
32. Hu, Albert S, Christine A Parlour, and Uday Rajan, 2019, Cryptocurrencies: Stylized facts on a new investible instrument, Financial Management 48, 1049–1068.
33. Borri, Nicola, 2019, Conditional tail-risk in cryptocurrency markets, Journal of Empirical Finance 50, 1–19.
34. Griffin, John M, and Amin Shams, 2020, Is Bitcoin really untethered?, The Journal of Finance 75, 1913–1964.

35. Corbet, Shaen, Brian Lucey, Andrew Urquhart, and Larisa Yarovaya, 2019, Cryptocurrencies as a financial asset: A systematic analysis, International Review of Financial Analysis 62, 182–199.
36. Choi, Sangyup, and Junhyeok Shin, 2022, Bitcoin: An inflation hedge but not a safe haven, Finance Research Letters 46, 102379.
37. Glaser, Florian, Kai Zimmermann, Martin Haferkorn, Moritz Christian Weber, and Michael Siering, 2014, Bitcoin-asset or currency? revealing users' hidden intentions, *Revealing Users' Hidden Intentions (April 15, 2014). ECIS*.
38. Eliason, Nat, 2021, Tokenomics 101: The Basics of Evaluating Cryptocurrencies,
39. Shiller, Robert J, 2019, Narrative economics: How stories go viral and drive major economic events, The Quarterly Journal of Austrian Economics 22, 620–627.
40. Breiman, Leo, 2001, Random forests, Machine learning 45, 5–32.