Replicating the Methodology of "Common Risk Factors in Cryptocurrency": A Python Approach

By Marc Gehring under the supervision of Professor Adrien d'Avernas

1. Introduction

The cryptocurrency market has experienced significant growth in recent years, attracting both academic and practical attention. In response, numerous studies have been conducted to investigate the factors driving the market and associated risks. Among these studies, "Common Risk Factors in Cryptocurrency" by Yukun Liu, Aleh Tsyvinski, and Xi Wu (2022) stands out as a comprehensive and influential work in the field.[1] In their paper, the authors find that three factors – cryptocurrency market, size, and momentum – are the primary drivers of cross-sectional expected cryptocurrency returns. They use a extensive list of price- and market-related return predictors, which have previously led to important results for equities, and find that ten cryptocurrency characteristics can be used to form successful long-short investment strategies. The authors then show that these investment strategies are explained by the cryptocurrency three-factor model.

This thesis aims to closely replicate the methodology used in "Common Risk Factors in Cryptocurrency" by implementing it in python code as a step-by-step guide with the goal of verifying the robustness of the findings and making the methodology more accessible and replicable for future researchers.[2] Throughout the thesis, the original methodology is thoroughly examined and any limitations and areas for improvement are identified.

The following section provides a comprehensive overview of the methodology, detailing the data sources, variables, and statistical methods employed, and compares the findings of the original paper with the new results. Subsequently, the thesis concludes with a discussion of the implications and recommendations for future research.

[1] We henceforth refer to this paper as the “original paper”.

[2] We confine our analysis to chapters II and III of the original paper, excluding the principal component analysis.

2. Methodology and comparison

In this section, we outline our efforts to replicate the methodology from the original paper using python. We detail the steps taken to produce each table and figure, then briefly compare our results to those in the original paper, though our primary goal is to replicate the methodology. The accompanying python file can be found at github.com/MarcGehring97/Crypto-Currency-Asset-Pricing.

In this thesis, we utilize a Jupyter notebook as the platform for our main analysis, taking advantage of its user-friendly interface and seamless integration of code, text, and visualizations. The structure of the notebook closely follows the sections of the original paper, starting with chapter *I. Data*. In this section, the researcher can filter coins based on criteria, such as market capitalization, and modify any additional assumptions made. Chapter *II. Cross-Sectional Return Predictors* involves the identification of long-short investment strategies that generate positive, statistically significant returns. Finally, a small number of explanatory factors for these strategies are investigated in chapter *III. Cryptocurrency Factors*. The goal of each section is to match the relevant tables from the original paper. Since direct LaTeX tabular format rendering is not possible in a Jupyter notebook, we created a PDF file to display the tables below each block of code. It is important to note that this file is overwritten each time a new table is rendered and must be copied if the researcher intends to save it.

The study begins by setting the start date to January 1st, 2014, as prior to this time, Bitcoin was the only actively traded cryptocurrency. The end date is set to July 31st, 2020 to make the results comparable with the original paper. The researcher is expected to specify the storage directory for the data. The code checks for the presence of relevant data files and skips the data generation section if these files already exist. The retrieval of the data involves obtaining the main data set, for which the authors of the original paper used CoinMarketCap (coinmarketcap.com). CoinMarketCap might be considered the most popular crypto data platform, but it does not provide access to historical data as part of its free plan. In this study, we propose the utilization of CoinGecko (coingecko.com) as an alternative data source. CoinGecko is the second largest crypto data platform and offers a free API for data gathering. However, it should be noted that the API has limitations, such as only allowing 50 calls per minute. The code for retrieving the data from CoinGecko is available in the *coingecko_data.py* file located in the *data_retrieval* folder. This folder also contains additional files that allow for the retrieval of data from other sources. We have decided to include additional data sources to enable researchers to explore additional explanatory factors. Given the API traffic limitation, it is advisable to download the data sets in smaller chunks to avoid having to restart the entire process in case of interruption. The code provides functionality for this purpose. Following the methodology described in the original paper, we transform the daily frequency of the data set to a weekly frequency by taking the last available data point each week, with a week being defined as starting on Monday and ending on Sunday. To account for potential artificial or erroneous returns, all returns greater than 80% in absolute value were set to missing. This threshold can be considered generous, and other researchers may adjust it to their preference. The data points where the market capitalization was below $1 million were filtered out, in line with the exclusion principle proposed in the original paper. However, all values for Bitcoin, Ripple, and Ethereum are still retained since these cryptocurrencies were considered the most important ones in the early days of public cryptocurrency trading. Moreover, we impute all missing values by the mean of the respective coin's return series, but only consider data after the first non-missing value. We opted for this method for its simplicity and since it can effectively fill in missing data points. Mean imputation can also help to preserve the sample size and distribution of the data. This is important for maintaining the validity of the analysis (Schafer, 1997). For the market capitalization, we impute all missing values after the first non-missing value by the average of the last and next available data values, a method known as linear interpolation. By using the average of the surrounding values, the imputed value is likely to be representative of the underlying trends in the data, rather than introducing an outlier. However, this method assumes a linear relationship between the missing value and the surrounding values, which may not always be accurate (Little and Rubin, 2002).[3] Next, the risk-free rate is retrieved as the one-month Treasury bill rate from French's website (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french). In the original paper, the authors are using the same rate. The code for this can be found in the *stock_factors_data.py* file located in the *data_retrieval* folder. For missing values that occurred on weekends, we impute the average of the values of Friday and Monday. The remaining missing data points are imputed using linear interpolation by taking the

[3] Unfortunately, the authors of the original paper do not provide an explanation for how they dealt with missing values. They neither include an explanation in their internet appendix.

average of the last and next available data values. After preprocessing the data, we compute the weekly value-weighted coin market return series and the coin market excess return series using the risk-free rate. We reindex the dataframes to the full date range after each operation to ensure consistency in the dimensionality of the data. Finally, the summary statistics for the daily and weekly return series are presented in Table I. Comparing this to the matching table in the original paper, we can see that all numbers differ substantially. This divergence can mainly be explained by the different data source, the possibly differing filtering process, and the assumptions made regarding the frequency conversion. In particular, we can see that the number of coins differs from the data in the original paper every year. This could mean that the data set we use contains fewer coins or that we are applying a finer filtering procedure.

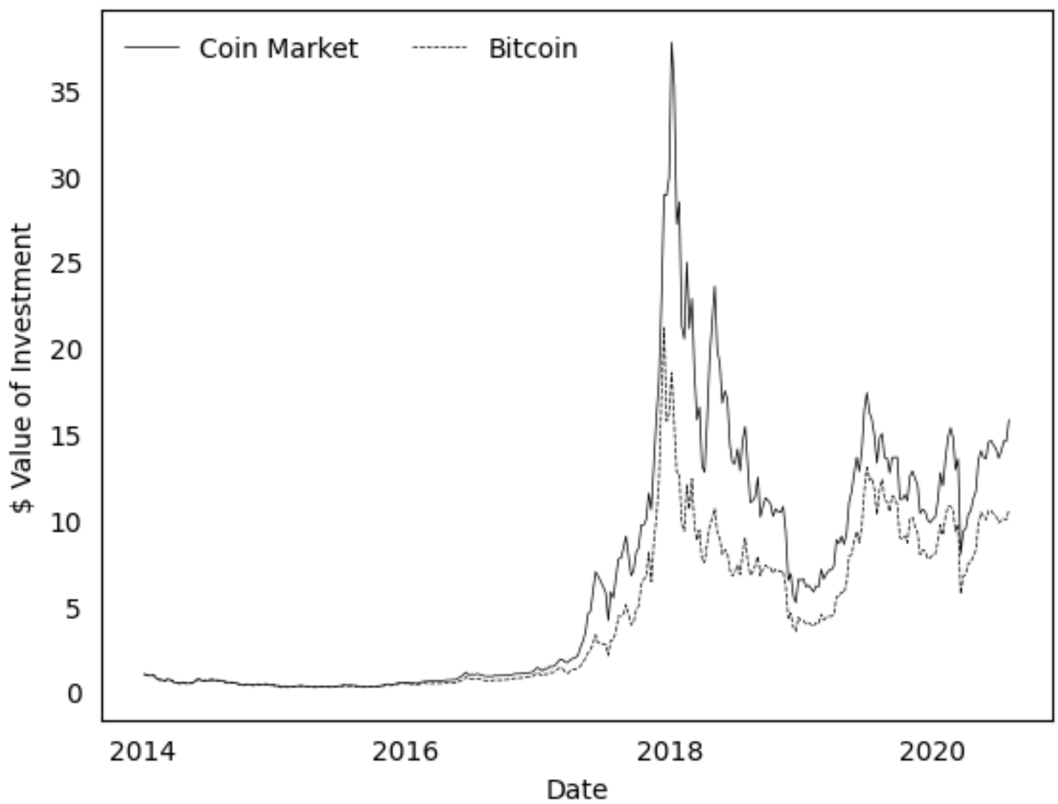
Table I
Summary Statistics

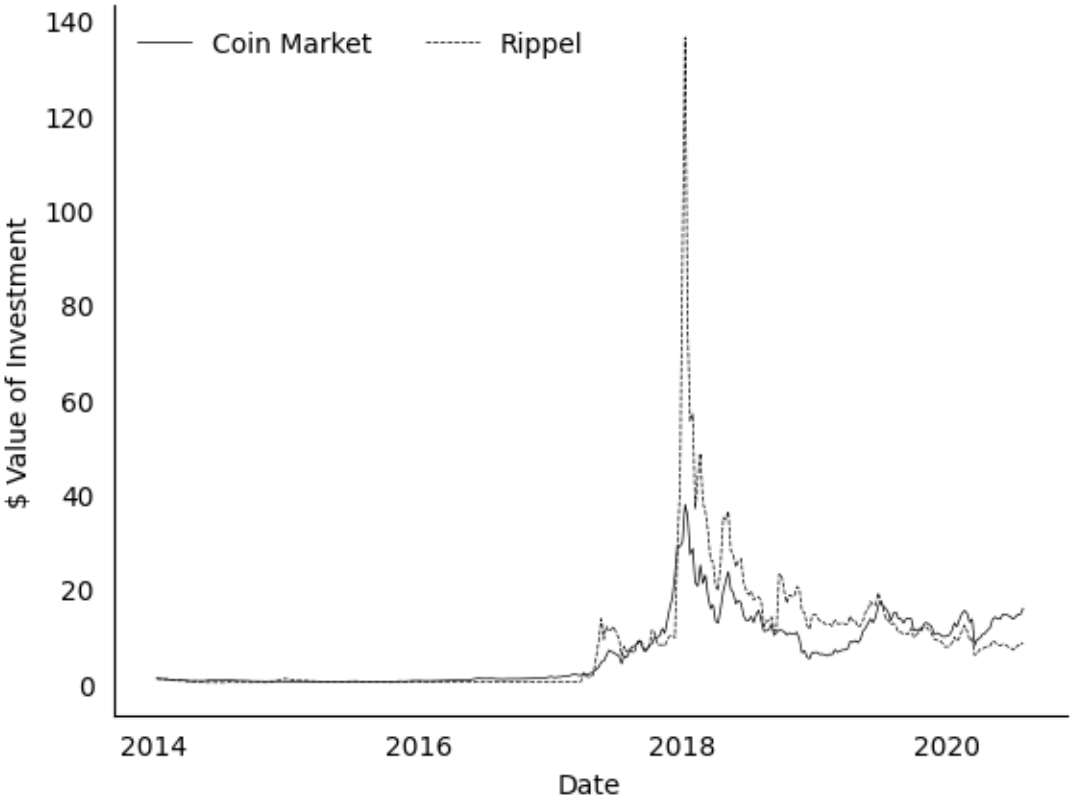
Panel A reports the number of coins, the mean and median of market capitalization, and the mean and median of daily trading price volume by year. Panel B reports the characteristics of coin market index returns, Bitcoin returns, Ripple returns, and Ethereum returns. The coin market index returns, Bitcoin returns, and Ripple returns start from the first week of 2014. The Ethereum returns start from the 32^{nd} week of 2015.

Panel A. Characteristics by Year					
		Market Cap (mil)		Volume (thous)	
Year	Number	Mean	Median	Mean	Median
2014	63	100.21	0.27	2,452.33	42.18
2015	42	36.28	0.12	13,389.12	16.28
2016	91	70.93	0.30	43,115.99	24.43
2017	386	554.03	5.80	25,091.91	243.43
2018	891	414.55	6.75	23,510.86	255.38
2019	888	220.53	2.01	121,192.48	304.53
2020	819	237.15	1.39	204,838.37	434.33
Full	12,468	281.48	2.07	91,964.79	263.86

Panel B. Return Characteristics					
	Mean	Median	*SD*	Skewness	Kurtosis
Coin Market Return	0.013	0.012	0.103	0.019	1.558
Bitcoin Return	0.013	0.007	0.107	0.177	1.750
Ripple Return	0.037	0.013	0.188	1.329	3.762
Ethereum Return	0.033	-0.009	0.341	8.922	111.075

We use the next block of code to visualize the return indices of the market return series and the return series of the major cryptocurrencies Bitcoin, Ripple, and Ethereum in Figure I. The indices are normalized to a value of 1 on the 1st of January, 2014. In the event of missing data, the last observed value is employed for imputation purposes. As mentioned earlier, we adjusted the filtering threshold for excessively high and low returns to more closely replicate the coin market return trajectory from the original paper. Interestingly, though, the Ethereum return series exhibits a far larger series maximum at about 1,000 times compared to 500 times in the original data. Still, the overall shape of the graphs looks almost identical. Apparently, the two data sources differ in a few Ethereum data points that make the series jump higher in this case. The Bitcoin and Ripple return series seem to resemble the ones in the original paper in both shape and magnitude.





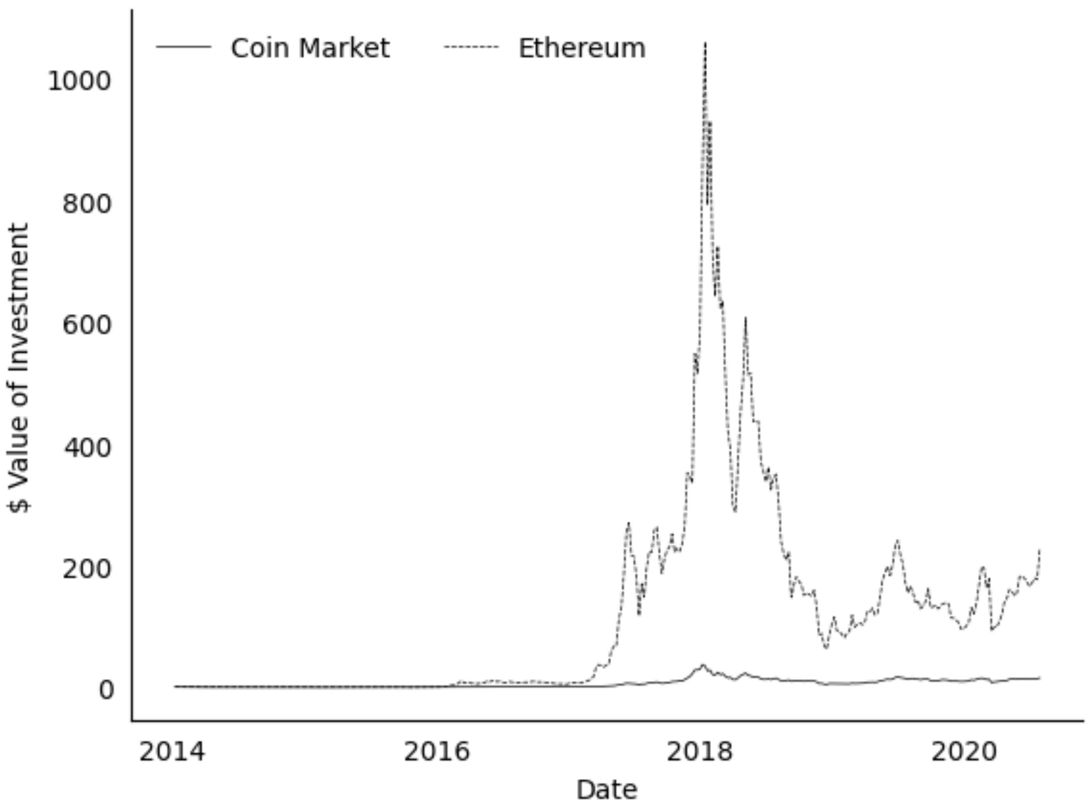


Figure 1. Cryptocurrency market and major coins. This figure plots the aggregate cryptocurrency market against Bitcoin, Ripple, and Ethereum.

In the following section of the notebook, the computation of the five quintile excess return series as well as the long-short investment strategies for different cryptocurrency trading variables is performed. Section A focuses on computing these returns for size characteristics such as *MCAP* (log last-day market capitalization in the portfolio formation week), *PRC* (log last-day price in the portfolio formation week), *MAXDPRC* (the maximum price of the portfolio formation week), and *AGE* (number of days listed since the time period began on January 1st, 2014). The quintile return series are computed by dividing all coins into quintiles based on each of the aforementioned characteristics on a weekly basis and calculating the value-weighted return of each quintile in the following week. Subsequently, the risk-free rate is subtracted from the individual quintile return series for each characteristic. The results of the size strategy returns are presented in Table II. The results agree with the original paper regarding the market capitalization factor but unlike the original paper deliver statistical insignificance for the other three factors.

Table II
Size Strategy Returns

This table reports the mean quintile portfolio returns based on the market capitalization, last-day price, and maximum day price measures. The mean returns are the time-series averages of weekly value-weighted portfolio excess returns. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Quintiles					
	1	2	3	4	5	5-1
MCAP	**Low**				**High**	
Mean	0.047***	0.010	0.012*	0.008	0.010*	-0.037***
t(Mean)	(7.94)	(1.52)	(1.80)	(1.30)	(1.83)	(-7.36)
PRC	**Low**				**High**	
Mean	0.007	0.012	0.003	-0.003	0.011*	0.004
t(Mean)	(0.96)	(1.52)	(0.37)	(-0.38)	(1.91)	(0.66)
MAXDPRC	**Low**				**High**	
Mean	0.007	0.011	0.001	0.000	0.011*	0.003
t(Mean)	(1.04)	(1.32)	(0.14)	(0.04)	(1.90)	(0.56)
AGE	**Low**				**High**	
Mean	0.002	0.010	0.011	0.030***	0.010*	0.008
t(Mean)	(0.24)	(1.15)	(1.42)	(2.91)	(1.69)	(1.19)

Additionally, in section B we also calculate the quintile return series for the momentum characteristics, including *r 1,0* (past one-week return), *r 2,0* (past two-week return), *r 3,0* (past three-week return), *r 4,0* (past four-week return), *r 4,1* (past one-to-four-week return), *r 8,0* (past eight-week return), *r 16,0* (past 16-week return), *r 50,0* (past 50-week return), and *r 100,0* (past 100-week return). The results are shown in Table III. Here, we observe similarities to the original paper in terms of the signs and magnitudes of the quintile returns and the long-short investment strategies. Strikingly, we obtain the tendency that longer-term momentum strategies are less likely to yield statistically significant long-short returns.

Table III
Momentum Strategy Returns

This table reports the mean quintile portfolio returns based on the past one-week, two-week, three- week, four-week, and one-to-four-week return measures. The mean returns are the time-series averages of weekly value-weighted portfolio excess returns. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Quintiles					
	1	2	3	4	5	5-1
r 1,0	**Low**				**High**	
Mean	-0.003	0.003	0.006	0.011*	0.020**	0.022***
t(Mean)	(-0.37)	(0.43)	(0.92)	(1.73)	(2.52)	(3.10)
r 2,0	**Low**				**High**	
Mean	0.001	-0.006	0.004	0.013*	0.025***	0.023***
t(Mean)	(0.17)	(-0.84)	(0.58)	(1.83)	(3.09)	(3.11)
r 3,0	**Low**				**High**	
Mean	0.000	-0.009	0.002	0.012*	0.029***	0.029***
t(Mean)	(0.03)	(-1.31)	(0.34)	(1.74)	(3.49)	(3.46)
r 4,0	**Low**				**High**	
Mean	0.000	0.001	0.001	0.007	0.028***	0.027***
t(Mean)	(0.06)	(0.08)	(0.21)	(1.10)	(3.44)	(3.42)
r 4,1	**Low**				**High**	
Mean	-0.001	-0.004	0.005	0.013*	0.021***	0.023***
t(Mean)	(-0.18)	(-0.55)	(0.70)	(1.93)	(2.67)	(2.96)
r 8,0	**Low**				**High**	
Mean	0.012	0.006	0.003	0.007	0.026***	0.014*
t(Mean)	(1.46)	(0.78)	(0.49)	(1.04)	(3.39)	(1.79)

	Quintiles					
	1	2	3	4	5	5-1
r 8,0	**Low**				**High**	
Mean	0.012	0.006	0.003	0.007	0.026***	0.014*
t(Mean)	(1.46)	(0.78)	(0.49)	(1.04)	(3.39)	(1.79)
r 16,0	**Low**				**High**	
Mean	0.014*	0.011	0.009	0.011	0.021***	0.007
t(Mean)	(1.66)	(1.50)	(1.25)	(1.49)	(2.76)	(0.90)
r 50,0	**Low**				**High**	
Mean	0.017*	-0.002	0.011	0.009	0.011*	-0.006
t(Mean)	(1.96)	(-0.20)	(1.45)	(1.36)	(1.68)	(-0.82)
r 100,0	**Low**				**High**	
Mean	0.008	0.004	0.012	0.004	0.015**	0.007
t(Mean)	(0.87)	(0.48)	(1.41)	(0.56)	(2.05)	(0.84)

For the volume characteristics in section C, we consider the predictors *VOL* (log average daily volume in the portfolio formation week), *PRCVOL* (log average daily volume times price in the portfolio formation week), and *VOLSCALED* (log average daily volume times price scaled by market capitalization in the portfolio formation week). The results of this analysis are depicted in Table IV. Contrary to the original paper, *PRCVOL* is not statistically significant, though the individual quintiles show signs of statistical significance.

Table IV
Volume Strategy Returns

This table reports the mean quintile portfolio returns based on the price volume measure. The mean returns are the time-series averages of weekly value-weighted portfolio excess returns. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Quintiles					
	1	2	3	4	5	5-1
VOL	**Low**				**High**	
Mean	0.017***	0.014**	0.007	0.012*	0.010*	-0.007
t(Mean)	(2.85)	(2.16)	(1.13)	(1.72)	(1.85)	(-1.43)

	Quintiles					
	1	2	3	4	5	5-1
PRCVOL	**Low**				**High**	
Mean	0.009	0.012	0.003	-0.000	0.011*	0.002
t(Mean)	(1.19)	(1.47)	(0.45)	(-0.02)	(1.88)	(0.33)
VOLSCALED	**Low**				**High**	
Mean	0.007	0.000	0.017*	0.011	0.013*	0.006
t(Mean)	(0.94)	(0.03)	(1.96)	(1.61)	(1.90)	(0.80)

Finally, we examine the quintile return series for the volatility characteristics in section D. The predictors for this analysis include *BETA* (The regression coefficient ${\beta^i}_{CMKT}$ in$R_i - R_f = \alpha^i + {\beta^i}_{CMKT} CMKT + \epsilon_i$ using daily returns of the previous 365 days before the formation week), *BETA2* (*BETA* squared), *IDIOVOL* (idiosyncratic volatility, measured as the standard deviation of the residual after estimating $R_i - R_f = \alpha^i + {\beta^i}_{CMKT} CMKT + \epsilon_i$ using daily returns of the previous 365 days before the formation week), *RETVOL* (standard deviation of daily returns in the portfolio formation week), *MAXRET* (maximum daily return of the portfolio formation week), *DELAY* (the improvement in R^2 in $R_i - R_f = \alpha^i + {\beta^i}_{CMKT} CMKT + {\beta^i}_{CMKT_{-1}} CMKT_{-1} + {\beta^i}_{CMKT_{-2}} CMKT_{-2} + \epsilon_i$ where $CMKT_{-1}$ and $CMKT_{-2}$ are the lagged one- and two-day coin market index excess returns, compared to using only current coin market excess returns using daily returns of the previous 365 days before the formation week), *STDPRCVOL* (log standard deviation of price volume in the portfolio formation week), and *DAMIHUD* (average absolute daily return divided by price volume in the portfolio formation week). The results of this analysis are presented in Table V. Unlike the authors of the original paper, we do not obtain statistical significance for the *STDPRCVOL* factor but find that the *BETA* factor is statistically significant at the 10% significance level.

Table V
Volatility Strategy Returns

This table reports the mean quintile portfolio returns based on the standard deviation of price volume. The mean returns are the time-series averages of weekly value-weighted portfolio excess returns. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Quintiles					
	1	2	3	4	5	5-1
BETA	**Low**				**High**	
Mean	-0.006	0.001	0.016*	0.009	0.016**	0.022***
t(Mean)	(-0.84)	(0.08)	(1.78)	(1.12)	(2.56)	(3.20)
BETA2	**Low**				**High**	
Mean	0.000	-0.002	0.014	0.012	0.015**	0.015**
t(Mean)	(0.01)	(-0.28)	(1.59)	(1.59)	(2.25)	(2.13)
IDIOVOL	**Low**				**High**	
Mean	0.013*	0.014*	0.010	0.004	-0.002	-0.015**
t(Mean)	(1.93)	(1.79)	(1.19)	(0.63)	(-0.28)	(-2.06)
RETVOL	**Low**				**High**	
Mean	0.008	0.005	0.012	0.011	0.005	-0.003
t(Mean)	(1.34)	(0.65)	(1.58)	(1.35)	(0.64)	(-0.32)
MAXRET	**Low**				**High**	
Mean	0.000	0.010	0.006	0.012	0.008	0.008
t(Mean)	(0.05)	(1.50)	(0.85)	(1.54)	(0.96)	(0.99)
DELAY	**Low**				**High**	
Mean	0.010	0.013	0.007	0.013*	0.012	0.002
t(Mean)	(1.52)	(1.65)	(1.08)	(1.80)	(1.45)	(0.31)
STDPRCVOL	**Low**				**High**	
Mean	0.011*	0.014**	0.011	0.010	0.011*	-0.001
t(Mean)	(1.77)	(2.18)	(1.64)	(1.58)	(1.90)	(-0.13)
DAMIHUD	**Low**				**High**	
Mean	0.011*	0.005	0.010	0.010	0.016**	0.005
t(Mean)	(1.87)	(0.72)	(1.46)	(1.58)	(2.40)	(0.99)

The subsequent chapter of the analysis investigates whether a limited number of factors can explain the long-short investment strategies identified previously. To this end, a one-factor model is run using the cryptocurrency market excess return (*CMKT*), also referred to as the cryptocurrency Capital Asset Pricing Model (CAPM). The dependent variables in this analysis are the various long-short investment strategies minus the risk-free rate. The results of the one-factor model are presented in Table VI. Notably, we receive similar results for the alphas of the momentum strategies in terms of their signs, magnitudes, and levels of significance. Yet, the results for the other factors converge widely from the ones obtained in the original paper.

Table VI
Cryptocurrency One-Factor Model

This table reports results for the cryptocurrency one-factor model adjustment of the 10 successful long-short strategies. The pricing model is

$$R_i - R_f = \alpha^i + \beta^i_{CMKT} CMKT + \epsilon_i,$$

where $CMKT$ is the cryptocurrency excess market return. t-Statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels. m.a.e. and $\overline{R^2}$ are the mean absolute pricing error and the average R^2 of the five portfolios, respectively.

	α	$t(\alpha)$	β_{CMKT}	$t(\beta_{CMKT})$	R^2	m.a.e.	$\overline{R^2}$
MCAP	-0.044***	(-9.42)	0.340***	(7.62)	0.146	0.062	0.609
PRC	-0.002	(-0.26)	0.214***	(3.78)	0.040	0.074	0.567
MAXDPRC	-0.002	(-0.36)	0.211***	(3.72)	0.039	0.075	0.561
AGE	0.002	(0.38)	0.194***	(3.12)	0.028	0.080	0.544
r 1,0	0.018**	(2.50)	0.090	(1.30)	0.005	0.093	0.518
r 2,0	0.019**	(2.55)	0.081	(1.11)	0.004	0.099	0.529
r 3,0	0.025***	(2.91)	0.123	(1.52)	0.007	0.110	0.519
r 4,0	0.023***	(2.85)	0.104	(1.36)	0.005	0.103	0.537
r 4,1	0.019**	(2.50)	0.017	(0.22)	0.000	0.099	0.526
r 8,0	0.010	(1.28)	0.063	(0.80)	0.002	0.103	0.537
r 16,0	0.003	(0.32)	0.099	(1.27)	0.005	0.099	0.519
r 50,0	-0.009	(-1.28)	-0.028	(-0.42)	0.001	0.084	0.606
r 100,0	0.003	(0.43)	-0.086	(-1.19)	0.006	0.086	0.569
VOL	-0.013***	(-2.86)	0.282***	(6.41)	0.108	0.059	0.609
PRCVOL	-0.003	(-0.57)	0.213***	(3.68)	0.038	0.077	0.562
VOLSCALED	0.002	(0.23)	0.095	(1.33)	0.005	0.090	0.535
BETA	0.011*	(1.75)	0.457***	(7.67)	0.170	0.064	0.535
BETA2	0.006	(0.87)	0.327***	(5.22)	0.086	0.071	0.541
IDIOVOL	-0.008	(-1.43)	-0.656***	(-11.72)	0.323	0.059	0.509
RETVOL	-0.005	(-0.57)	-0.115	(-1.46)	0.006	0.106	0.527
MAXRET	0.004	(0.54)	0.038	(0.48)	0.001	0.103	0.544
DELAY	-0.001	(-0.15)	-0.045	(-0.68)	0.002	0.077	0.552
STDPRCVOL	-0.007	(-1.34)	0.274***	(5.62)	0.085	0.062	0.618
DAMIHUD	0.005	(0.93)	-0.262***	(-5.11)	0.071	0.068	0.583

We finally set up three multi-factor models, adding the size and the momentum factors. The size factor, *CSMB*, is determined by comparing the returns of small and large portfolios. The *CMOM* is established by utilizing three-week momentum and forming portfolios by combining 2 sets of 3 portfolios. Each week, coins are divided into two groups based on their size, then further separated into three portfolios based on their past three-week performance. These portfolios consist of the bottom 30%, middle 40%, and top 30% of coins ranked by their past returns. The momentum factor is then constructed as

$$CMOM = \frac{1}{2}(Small\ High + Big\ High) + \frac{1}{2}(Small\ Low + Big\ Low).$$

Model (1) incorporates the *CMKT* and *CSMB* factor, Model (2) includes the *CMKT* and momentum factor, and Model (3) includes all three factors, i.e. *CMKT*, *CSMB*, and *CMOM*. The results of these models are reported in Table VII, including the model statistics. Comparing these results to the findings of the original paper, we notice again that there appears to be most agreement for the momentum factors. All other factors differ in terms of sign, magnitude, and statistical significance from the original paper.

Table VII
Cryptocurrency Factor Models

This table reports results on the cryptocurrency factor adjustments of the 10 successful long-short strategies. $CMKT$ is the cryptocurrency excess market return, $CSMB$ is the cryptocurrency size factor, and $CMOM$ is the cryptocurrency momentum factor. $t-Statistics$ are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels. m.a.e. and $\overline{R^2}$ are the mean absolute pricing error and the average R^2 of the five portfolios, respectively.

		Cons	t	CMKT	t	CSMB	t	CMOM	t	R^2	m.a.e.	$\overline{R^2}$
MCAP	(1)	-0.047***	(-10.07)	0.047	(0.49)	0.047	(0.49)			0.174	0.061	0.620
MCAP	(2)	-0.050***	(-11.09)	0.301***	(7.04)			0.301	(7.04)	0.233	0.059	0.635
MCAP	(3)	-0.053***	(-11.64)	0.044	(0.47)	0.044	(0.47)	0.339	(3.12)	0.254	0.058	0.645
PRC	(1)	-0.006	(-1.04)	-0.173	(-1.40)	-0.173	(-1.40)			0.074	0.073	0.575
PRC	(2)	0.001	(0.22)	0.233***	(4.09)			0.233	(4.09)	0.054	0.074	0.571
PRC	(3)	-0.003	(-0.54)	-0.172	(-1.40)	-0.172	(-1.40)	0.534	(3.70)	0.091	0.072	0.580
MAXDPRC	(1)	-0.007	(-1.17)	-0.191	(-1.55)	-0.191	(-1.55)			0.075	0.073	0.569
MAXDPRC	(2)	0.001	(0.11)	0.230***	(4.02)			0.230	(4.02)	0.052	0.074	0.565
MAXDPRC	(3)	-0.004	(-0.67)	-0.190	(-1.55)	-0.190	(-1.55)	0.554	(3.84)	0.092	0.072	0.574
AGE	(1)	-0.001	(-0.21)	-0.130	(-0.95)	-0.130	(-0.95)			0.048	0.078	0.551
AGE	(2)	0.005	(0.69)	0.208***	(3.31)			0.208	(3.31)	0.034	0.080	0.549
AGE	(3)	0.001	(0.11)	-0.129	(-0.95)	-0.129	(-0.95)	0.445	(2.77)	0.056	0.079	0.557
r 1,0	(1)	0.016**	(2.15)	-0.092	(-0.60)	-0.092	(-0.60)			0.010	0.093	0.521
r 1,0	(2)	0.014*	(1.86)	0.054	(0.78)			0.054	(0.78)	0.034	0.093	0.529
r 1,0	(3)	0.012	(1.60)	-0.092	(-0.61)	-0.092	(-0.61)	0.193	(1.08)	0.037	0.092	0.534
r 2,0	(1)	0.018**	(2.33)	-0.020	(-0.12)	-0.020	(-0.12)			0.005	0.099	0.533
r 2,0	(2)	0.013*	(1.71)	0.034	(0.47)			0.034	(0.47)	0.051	0.098	0.542
r 2,0	(3)	0.012	(1.59)	-0.021	(-0.14)	-0.021	(-0.14)	0.073	(0.39)	0.051	0.098	0.546
r 3,0	(1)	0.023***	(2.68)	0.013	(0.07)	0.013	(0.07)			0.008	0.109	0.523
r 3,0	(2)	0.018**	(2.12)	0.075	(0.93)			0.075	(0.93)	0.048	0.108	0.534
r 3,0	(3)	0.017**	(1.99)	0.011	(0.06)	0.011	(0.06)	0.084	(0.40)	0.048	0.108	0.538
r 4,0	(1)	0.019**	(2.34)	-0.203	(-1.21)	-0.203	(-1.21)			0.018	0.103	0.543
r 4,0	(2)	0.014*	(1.80)	0.041	(0.55)			0.041	(0.55)	0.083	0.101	0.552
r 4,0	(3)	0.011	(1.42)	-0.206	(-1.27)	-0.206	(-1.27)	0.325	(1.71)	0.091	0.100	0.558
r 4,1	(1)	0.018**	(2.32)	-0.055	(-0.34)	-0.055	(-0.34)			0.001	0.098	0.529
r 4,1	(2)	0.015*	(1.90)	-0.015	(-0.21)			-0.015	(-0.21)	0.022	0.100	0.537
r 4,1	(3)	0.014*	(1.80)	-0.057	(-0.35)	-0.057	(-0.35)	0.054	(0.28)	0.022	0.099	0.541
r 8,0	(1)	0.007	(0.89)	-0.172	(-1.00)	-0.172	(-1.00)			0.009	0.102	0.544
r 8,0	(2)	0.005	(0.57)	0.018	(0.23)			0.018	(0.23)	0.037	0.102	0.547
r 8,0	(3)	0.002	(0.29)	-0.172	(-1.02)	-0.172	(-1.02)	0.253	(1.27)	0.042	0.101	0.554
r 16,0	(1)	-0.001	(-0.14)	-0.176	(-1.05)	-0.176	(-1.05)			0.015	0.097	0.526
r 16,0	(2)	-0.001	(-0.12)	0.072	(0.91)			0.072	(0.91)	0.021	0.099	0.527
r 16,0	(3)	-0.004	(-0.49)	-0.176	(-1.05)	-0.176	(-1.05)	0.333	(1.68)	0.029	0.097	0.535

		Cons	t	CMKT	t	CSMB	t	CMOM	t	R^2	m.a.e.	$\overline{R^2}$
r 50,0	(1)	-0.008	(-1.04)	0.059	(0.41)	0.059	(0.41)			0.002	0.084	0.607
r 50,0	(2)	-0.006	(-0.78)	-0.000	(-0.00)			-0.000	(-0.00)	0.026	0.085	0.609
r 50,0	(3)	-0.005	(-0.63)	0.058	(0.41)	0.058	(0.41)	-0.080	(-0.47)	0.027	0.085	0.611
r 100,0	(1)	0.003	(0.32)	-0.124	(-0.86)	-0.124	(-0.86)			0.006	0.086	0.574
r 100,0	(2)	0.007	(0.81)	-0.061	(-0.84)			-0.061	(-0.84)	0.019	0.086	0.571
r 100,0	(3)	0.005	(0.63)	-0.125	(-0.87)	-0.125	(-0.87)	0.093	(0.52)	0.020	0.086	0.576
VOL	(1)	-0.016***	(-3.34)	0.074	(0.76)	0.074	(0.76)			0.123	0.058	0.617
VOL	(2)	-0.015***	(-3.21)	0.272***	(6.14)			0.272	(6.14)	0.116	0.058	0.614
VOL	(3)	-0.017***	(-3.64)	0.072	(0.75)	0.072	(0.75)	0.264	(2.32)	0.130	0.058	0.622
PRCVOL	(1)	-0.008	(-1.33)	-0.173	(-1.37)	-0.173	(-1.37)			0.071	0.075	0.571
PRCVOL	(2)	-0.001	(-0.20)	0.227***	(3.89)			0.227	(3.89)	0.046	0.077	0.566
PRCVOL	(3)	-0.006	(-0.93)	-0.172	(-1.37)	-0.172	(-1.37)	0.528	(3.56)	0.081	0.075	0.576
VOLSCALED	(1)	-0.001	(-0.07)	-0.089	(-0.56)	-0.089	(-0.56)			0.010	0.089	0.539
VOLSCALED	(2)	0.008	(1.06)	0.137*	(1.93)			0.137	(1.93)	0.046	0.087	0.549
VOLSCALED	(3)	0.005	(0.71)	-0.086	(-0.55)	-0.086	(-0.55)	0.294	(1.61)	0.053	0.086	0.555
BETA	(1)	0.007	(1.11)	0.227*	(1.84)	0.227*	(1.84)			0.183	0.063	0.545
BETA	(2)	0.014**	(2.22)	0.481***	(8.05)			0.481	(8.05)	0.189	0.064	0.543
BETA	(3)	0.010	(1.54)	0.226*	(1.86)	0.226*	(1.86)	0.355*	(2.40)	0.205	0.063	0.553
BETA2	(1)	0.001	(0.21)	0.067	(0.52)	0.067	(0.52)			0.103	0.070	0.551
BETA2	(2)	0.008	(1.27)	0.348***	(5.53)			0.348	(5.53)	0.102	0.071	0.548
BETA2	(3)	0.004	(0.59)	0.066	(0.51)	0.066	(0.51)	0.393	(2.52)	0.121	0.069	0.558
IDIOVOL	(1)	-0.006	(-0.97)	-0.506***	(-4.35)	-0.506***	(-4.35)			0.328	0.059	0.521
IDIOVOL	(2)	-0.012**	(-2.06)	-0.686***	(-12.32)			-0.686	(-12.32)	0.350	0.059	0.515
IDIOVOL	(3)	-0.009	(-1.53)	-0.505***	(-4.43)	-0.505***	(-4.43)	-0.252***	(-1.82)	0.357	0.058	0.528
RETVOL	(1)	-0.004	(-0.48)	-0.061	(-0.35)	-0.061	(-0.35)			0.007	0.106	0.530
RETVOL	(2)	-0.008	(-0.93)	-0.126	(-1.58)			-0.126	(-1.58)	0.010	0.105	0.535
RETVOL	(3)	-0.007	(-0.82)	-0.065	(-0.37)	-0.065	(-0.37)	-0.080	(-0.39)	0.010	0.105	0.539
MAXRET	(1)	0.004	(0.42)	-0.034	(-0.20)	-0.034	(-0.20)			0.001	0.103	0.546
MAXRET	(2)	-0.001	(-0.12)	0.005	(0.07)			0.005	(0.07)	0.024	0.102	0.552
MAXRET	(3)	-0.002	(-0.18)	-0.038	(-0.22)	-0.038	(-0.22)	0.057	(0.28)	0.024	0.102	0.555
DELAY	(1)	0.002	(0.23)	0.114	(0.84)	0.114	(0.84)			0.008	0.077	0.557
DELAY	(2)	-0.004	(-0.55)	-0.067	(-1.02)			-0.067	(-1.02)	0.019	0.077	0.560
DELAY	(3)	-0.001	(-0.14)	0.115	(0.85)	0.115	(0.85)	-0.254	(-1.54)	0.027	0.076	0.566
STDPRCVOL	(1)	-0.010*	(-1.96)	0.001	(0.01)	0.001	(0.01)			0.107	0.061	0.628
STDPRCVOL	(2)	-0.009*	(-1.69)	0.262***	(5.31)			0.262	(5.31)	0.093	0.062	0.621
STDPRCVOL	(3)	-0.012**	(-2.24)	-0.001	(-0.01)	-0.001	(-0.01)	0.347	(2.75)	0.113	0.061	0.631
DAMIHUD	(1)	0.007	(1.26)	-0.099	(-0.87)	-0.099	(-0.87)			0.078	0.068	0.592
DAMIHUD	(2)	0.007	(1.35)	-0.248***	(-4.80)			-0.248	(-4.80)	0.081	0.067	0.588
DAMIHUD	(3)	0.009	(1.63)	-0.097	(-0.86)	-0.097	(-0.86)	-0.199	(-1.50)	0.088	0.067	0.596

3. Conclusion

To summarize, this thesis aimed to provide a tool kit for other academics to replicate the methodology used in the paper by Liu, Tsyvinski, and Wu (2022). While some results obtained in this thesis differ from the original paper's findings, the primary goal was to replicate the methodology. The differences can mainly be explained by the different data sources, the possibly differing filtering process, and the assumptions made regarding the frequency conversion. Nonetheless, the signs and significance levels of the long-short investment strategies tend to agree with the findings of the original paper, underlining the validity of the replication.

The tool kit is valuable for both academia and practitioners in the cryptocurrency industry and provides a strong foundation for future research, making it easier for others to replicate the results. Future research can expand the analysis to include more cryptocurrencies and a longer time frame to gain a more comprehensive understanding of the factors affecting the cryptocurrency market and their impact on expected returns. Furthermore, examining the interaction between these factors and other market variables, such as regulation and technology advancements, can add depth to our understanding of the cryptocurrency market and inform investment decisions.

References

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). New York, U.S.: John Wiley & Sons, Inc.

Liu, Y., Tsyvinski, A., and Wu, X. (2022). Common risk factors in cryptocurrency. *Journal of Finance*, 77, 1133-1177.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data* (1st ed.). New York, U.S.: Chapman and Hall/CRC.