Erasmus school of economics

An expansion of the three-factor model for cryptocurrencies

Thesis in Financial Econometrics

Student:

Leon Verweij

Student number:

541475

Supervisor: Terri van der Zwan

Second assessor: Rasmus Lönn

July 9, 2022

Abstract

Since the rise of cryptocurrencies, more research has focused on explaining cryptocurrency returns. A part of the research on stock returns is replicated on cryptocurrencies, like the factor models. The three-factor model that includes a market, size and momentum factor has been proven to have explanatory power on cryptocurrency returns. This research extends the three-factor model to increase the explanatory power by adding seven factors that individually have a significant effect on cryptocurrencies. Furthermore, it identifies which of the ten factors contribute the most to the model's performance. I use a data set of 2024 cryptocurrencies from 2014 until now. The ten-factor model outperforms the cryptocurrency version of the CAPM and the three-factor model. Furthermore, it can explain the cross-section of cryptocurrency for most periods when looking at a window of one year. Moreover, the factors that significantly contribute to the model are a volume and price factor, besides the factors from the three-factor model. The results of this paper can be useful for institutional investors that use cryptocurrencies in their trading strategies.

Contents

1 Introduction

Since the white paper by Nakamoto (2008), which established the model for a blockchain, thousands of cryptocurrencies have been developed, and billions of euros have been invested in these cryptocurrencies, especially by young people. Almost 50% of all cryptocurrency investors are between 25 and 34 years old (Jain et al., 2019). Lately, not only individuals are trading cryptocurrencies, but also institutional investors, e.g. hedge funds and investment banks, have started trading in cryptocurrencies (Steer, 2022). This raises the question whether there are ways to explain or forecast the returns of cryptocurrencies so that it is also profitable for these more prominent investors to put money in cryptocurrencies. An obvious way to do this is to study the existing methods that try to explain the returns of stocks.

Many studies have already researched asset pricing models, for example, factor models. The first model introduced by Sharpe (1964) and Lintner (1965) is the Capital Asset Pricing Model (CAPM). This model only includes one factor, which is the excess market return. The interpretation of the CAPM is straightforward, but it cannot explain all variations in stock returns. That is why there has been a lot of research to find other factors. As a result, the so-called 'factor-zoo' came into existence. This is a pool of hundreds of potential factors that drive stock returns, as described by Harvey et al. (2016) and Hou et al. (2020).

However, many of these factors cannot be constructed for the cryptocurrency market, since cryptocurrencies do not have financial and accounting data. Three factors that can be constructed for cryptocurrencies and have been investigated extensively are: the market, size and momentum factor. There is broad consensus about the effect of the market and size factor, but not all papers agree on the effect of the momentum factor. This may be due to the different periods they researched or because they used a different set of cryptocurrencies. Shahzad et al. (2021) and Shen et al. (2020) found that the one-week momentum factor generates significantly negative returns in the short term, which means that there is a reversal of payoffs. My paper also includes the size and reversal effect as factors in the factor model, plus the market factor, which is the excess market return. I add seven extra factors to increase the explanatory power of the model. These factors are proposed by Liu et al. (2022), who investigate the effect of 25 factors [1]. They find that seven other factors besides market capitalization and one-week momentum are significant. These factors are: price, maximum price, two-, three-, and four-week

[1] A list of these factors can be found in Appendix A

momentum, dollar volume and standard deviation of dollar volume. Together with the three earlier mentioned factors, these seven factors produce the ten-factor model.

However, these ten factors may not all be relevant when they are included in the model simultaneously. Therefore, I apply forward and backward selection to select the significant variables from the ten-factor model that contribute to the model's performance. To overcome the multicollinearity that results from highly correlated factors, I perform Principal Component Analysis (PCA) that constructs components that maximize the variance and that are uncorrelated. To counter overfitting, I reduce the number of factors by selecting the components that explain a large part of the variance. Furthermore, I use the Lasso by Tibshirani (1996) as a robustness check for the factor selection methods. To analyze the performance of the ten-factor model over time, I perform a rolling window regression.

My paper investigates five models: three-factor, ten-factor, PCA, forward selection and backward selection. These five models are evaluated based on their explanatory power, using the (adjusted) R^2 and the GRS test by Gibbons et al. (1989).

This paper contributes to the existing literature by extending the three-factor model with seven potentially significant factors. The goal of adding these seven extra factors is to test whether they can explain the cross-section of cryptocurrency returns. Furthermore, I search for the factors that have the most explanatory power. My research is performed on a database that includes 2024 cryptocurrencies from the beginning of 2014 until now.

I find that the reversal effect is present in my data set, so the returns of the cryptocurrencies tend to go back to the trend. This effect is larger for cryptocurrencies with a small market capitalization. The forward and backward selection methods both include a price factor and the standard deviation of the volume as relevant and significant factors. Lasso confirms this for the standard deviation factor. Furthermore, all models cannot fully explain the cryptocurrency returns over the whole period. However, a rolling window regression reveals that the ten-factor model can explain the returns in most periods.

The remaining of my paper is structured as follows. Section 2 goes more in-depth into the existing literature. Section 3 then describes the data used in this paper and how these data are transformed. After that, Section 4 explains the methods used. Section 5 shows and discusses the results. Finally, Section 6 draws conclusions and discusses the research.

2 Theory

As already mentioned, I start with a model that includes a market, size and momentum factor. The most well-known paper that investigates the effect of the market and size factor is by Fama and French (1993). They showed that excess market return, size of stocks and the book-to-market ratio are three risk factors that explain a part of the average stock returns. For cryptocurrencies there also is broad consensus about the effect of the market and size factor, e.g. Shahzad et al. (2021), Shen et al. (2020), Liu et al. (2020). They concluded that the market and size factor have significant explanatory power on the cryptocurrency returns. Furthermore, they found a negative relation between the returns and size, which means that cryptocurrencies with a small market capitalization tend to outperform those with a large market capitalization.

Momentum as a risk factor was first proposed by Jegadeesh and Titman (1993). It captures the effect that stocks tend to maintain price trends. This effect has been researched for many applications, e.g. in the equity markets (Rouwenhorst, 1998), commodities (Miffre and Rallis, 2007) and foreign exchange markets (Menkhoff et al., 2012). In all these applications, there is a significant positive momentum effect. Carhart (1997) adds the momentum factor to the three-factor model of Fama and French (1993) and also found significant positive effect. However, it is not clear what the effect of momentum is on cryptocurrencies. Tzouvanas et al. (2020) and Liu et al. (2020) found a positive effect of momentum on returns. However, other studies found contradicting results. Grobys and Sapkota (2019) found no evidence of significant momentum pay-offs but found that the effect is the other way around. Instead of the price maintaining its trend, it tends to go back. Shen et al. (2020) found the same result and called it the reversal factor, which means that the returns tend to reverse instead of maintaining their trend. These contradicting results show that further research is necessary to be able to conclude what the effect of momentum is on cryptocurrencies. Nonetheless, it could be that the momentum effect is time-varying or depends on the set of cryptocurrencies on which it is tested.

Recently, several other studies proposed new factors and models to explain cryptocurrency returns. For example, Shahzad et al. (2021) added the contagion risk factor to the size and momentum factor and found that this outperforms the three-factor model. Zhang and Li (2020) tested the effect of idiosyncratic volatility on cryptocurrency returns and found that they are positively related. Jia et al. (2021) added another two factors to the volatility factor, namely skewness and kurtosis. They also found a strong positive relation between volatility and cryptocurrency returns and between kurtosis and returns, and a negative relation between skewness and returns. Schwenkler and Zheng (2020) proposed a model that looked at the co-movement of cryptocurrencies and saw that returns can be predicted by looking at this co-movement. Besides research in modelling the returns, there has also been research in using cryptocurrencies in portfolios of conventional financial assets. Petukhina et al. (2020) showed that, depending on the investor's objectives, cryptocurrencies can improve the risk-return trade-off of portfolios. This means that cryptocurrencies can be a useful asset when diversifying portfolios.

The contribution of my paper to the abovementioned literature is that I add other significant factors to the three-factor model to increase the model's explanatory power. These new factors were investigated by Liu et al. (2022). They looked at many established factors in the cross-section of stock returns, which are compiled by Feng et al. (2020) and Chen and Zimmermann (2020). Liu et al. (2022) constructed the cryptocurrency versions of the factors for which it is possible to do so, resulting in 25 factors that could be used to explain the returns. The paper found that 9 of these 25 factors have significant explanatory power. These are: market capitalization, price, maximum price, one-, two-, three-, and four-week momentum, volume, and standard deviation of volume. The study also concluded that the size and momentum factor are important in capturing cross-sectional cryptocurrency returns. However, it did not investigate whether adding factors to the three-factor model increases the performance. That is why I investigate the possibility of a model that outperforms the three-factor model in explaining the cross-section of cryptocurrency returns. Moreover, I also investigate which of these ten factors contribute the most to the model's performance.

3 Data

This section first describes in Section 3.1 the cryptocurrencies used in my research. Section 3.2 determines the formation and holing period needed for the reversal factor. Thereafter, Section 3.3 shows how the factors are constructed and reports their descriptive statistics. Lastly, Section 3.4 describes the returns to be explained by the factors.

3.1 Data description

Several websites provide cryptocurrency data, e.g. CoinAPI, Cryptocompare, Coinmarketcap and Coingecko. I use the free API of Coingecko, a website that aggregates information from many exchanges and of most cryptocurrencies. Chuen et al. (2017), Lyandres et al. (2019) and Schwenkler and Zheng (2020) also used this API for their papers. Coingecko provides data on the price, market capitalization, trading volume and quantity in circulation for over 13,000 cryptocurrencies. I need the price, market capitalization and trading volume of the cryptocurrencies to construct the factors. I use the daily prices to construct weekly returns. The data starts on April 28th 2013 and is available until the present [2].

I select the cryptocurrencies that had their Initial Coin Offering (ICO) before March 31st 2019, such that I have enough data for each cryptocurrency. Another way to ensure I have enough data is by starting my research on April 3rd 2014. This results in a total of 2033 available cryptocurrencies over the whole period. These also include nine stablecoins, which are cryptocurrencies that are linked to a specific currency. That means they are supposed to remain around the same price level, so I exclude the stablecoins. A list of the stablecoins is included in Appendix B. Another criterion I implement is that I exclude pump-and-dump coins. Pump-and-dump coins generate really high returns in a short period because people manipulate the price, but shortly after that, the price collapses. I maintain this criterion by limiting the maximum return in a week to 200% and a minimum return of -200% [3]. So, I do not use the cryptocurrencies that exceed these returns in my research.

The number of active cryptocurrencies at a particular moment changes over time. Table 1 shows the distribution of the number of cryptocurrencies over the different years. It reports the mean of the number of cryptocurrencies during a specific year, which has grown monotonically from 2014 until 2019. However, after 2019, the number of cryptocurrencies decreases slightly. That is because the cryptocurrencies are selected based on their ICO being before 2019. After the beginning of 2019, some cryptocurrencies ceased to exist, so the number of cryptocurrencies decreased. Table 1 also shows the mean of the market capitalization and volume during the years. Both the market capitalization and volume have increased a lot since 2014. The mean of the market capitalization has grown from 128.78 to 1006.73 million dollars. The mean volume has grown from 895.25 to 41,106.67 thousand dollars.

[2] The data is accessed on June 12th 2022

[3] The results for the formation and holding period are robust to the less strict criterion of 300% and -300%

Table 1: Summary statistics

Year	Nr. of Coins	Market Cap (mil.)	Volume (thous.)
2014	70.53	128.78	895.25
2015	118.74	37.59	2331.55
2016	138.38	76.36	14599.29
2017	226.82	468.38	12994.65
2018	923.82	3824.41	18213.15
2019	1540.22	133.79	26881.96
2020	1486.25	202.77	45295.88
2021	1500.60	1048.91	90517.13
2022	2881.58	1006.73	43106.67

This table shows the mean of the number of coins, market capitalization and volume by year.

3.2 Formation and holding period

For the reversal factor, I need the length of the formation and holding period. The formation period is when the winners and losers are identified, and the holding period is how long there will be a position in these stocks. The winners are defined as the cryptocurrencies in the decile with the highest returns, and the losers are defined as the cryptocurrencies in the decile with the lowest returns. The formation and holding period are determined by constructing so-called J-K portfolios, in which the formation period is J weeks and the holding period is K weeks, as inspired by Jegadeesh and Titman (1993). For the formation and holding period, I consider a horizon of 1, 2, 3 and 4 weeks, which leads to 16 strategies in total. These strategies are used to construct three types of equally-weighted portfolios: buy, sell and buy-sell. The buy portfolio consists of the winners, the sell portfolio consists of the losers, and the buy-sell portfolio buys the winners and sells the losers and is a zero-cost portfolio.

Appendix C reports the mean returns from the 16 J-K strategies in combination with the three portfolios. All buy portfolios generate negative returns, with most of the returns being significant at a 1% level. Furthermore, all sell portfolios generate significant positive average returns at a significance level of 1%. Together, this results in only negative returns for the buy-sell portfolios at a significance level of 1%. Therefore, there is indeed a reversal effect, as in the paper of Shen et al. (2020). The 1-1 strategy, combined with the buy-sell portfolio, generates the most negative mean return (-0.154) and has the highest absolute t-statistic (21.373). Therefore, the best J-K strategy is the 1-1 strategy, so I use this for constructing the factors.

3.3 Factors

This section starts with a description of the factor construction in Section 3.3.1. Then in Section 3.3.2 I discuss the descriptive statistics of the factors. Lastly, I show the results of the PCA in Section 3.3.3

3.3.1 Construction of factors

For the market factor I need the market return, which is the excess market return. I compute the value-weighted market returns as:

$$MKT_t = \sum_{i=1}^{n} ret_{it} * \frac{cap_{it}}{totalCap_t}, \tag{1}$$

where MKT_t is the market return in week t, n is the total number of cryptocurrencies, ret_{it} is the return of cryptocurrency i in week t, cap_{it} is the market capitalization of cryptocurrency i in week t and $totalCap_t$ is the total market capitalization in week t, so $totalCap_t = \sum_{i=1}^{n} cap_{it}$. The market factor is computed as:

$$RMRF = MKT - Rf, \tag{2}$$

where Rf is the risk-free rate [4].

Secondly, for the size (SMB) factor, I sort the cryptocurrencies based on their market capitalization weekly. Following Fama and French (2012), the large-cap currencies are the currencies with the top 90% market capitalization, and the small-cap currencies are the currencies with the bottom 10% market capitalization. This results in two portfolios: the small-cap portfolio (S) and the large-cap portfolio (B).

Thirdly, I use the selected 1-1 strategy for the reversal (DMU) factor. The cryptocurrencies will be sorted based on their one-week prior returns. Again following Fama and French (2012), the breakpoints are the 30^{th} and 70^{th} percentiles, resulting in three portfolios: high prior returns (U), medium prior returns (M) and low prior returns (D)

[4]This is the one-month Treasury bill rate, downloaded from `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

The intersection of these two size portfolios and the three prior return portfolios results in six portfolios: SU, SM, SD, BU, BM, and BD. I compute the SMB factor as the difference between the equally weighted mean of the small-cap portfolios and the equally weighted mean of the large-cap portfolios:

$$SMB = 1/3(SU + SM + SD) - 1/3(BU + BM + BD). \tag{3}$$

I compute the DMU factor as the difference between the equally weighted mean of the low prior return portfolios and the equally weighted mean of the high prior return portfolios:

$$DMU = 1/2(SD + BD) - 1/2(SU + BU). \tag{4}$$

The definitions of the seven factors proposed by Liu et al. (2022) are reported in 2. These seven factors are: PRC, MAXPRC, MOM2, MOM3, MOM4, VOL and STDVOL.

Table 2: Factor definitions

Factor	Definition
PRC	The last day price in the portfolio formation period
MAXPRC	The maximum price of the portfolio formation period
MOM2	The two-week momentum
MOM3	The three-week momentum
MOM4	The four-week momentum
VOL	The average daily volume in the portfolio formation period
STDVOL	The standard deviation of volume in the portfolio formation period

Each week, I sort the cryptocurrencies into quintiles based on the abovementioned factors. The seven factors are constructed as the returns in the following week of the fifth quintile minus the returns of the first quintile.

3.3.2 Descriptive statistics of the factors

The descriptive statistics of all ten factors and their correlations are included in Appendix D. The absolute average return on all factors is higher than on RMRF. The fact that some average returns of the seven extra factors are negative means that the first quintile produces higher returns than the fifth quintile. So, to get positive returns, the fifth quintile should be subtracted from the first quintile. However, this does not change the explanatory power of a factor. Furthermore, all factors have a leptokurtic distribution. The absolute skewness for all factors except for RMRF is between 0.2 and 0.9. As expected, some correlations are very high. This holds for the correlations between the momentum factors, the correlation between the price factors and the correlation between the volume factors. As a result, there is a lot of multicollinearity [5].

When there is multicollinearity, the coefficient estimates are not stable. Unstable coefficients means that the t-statistics and p-values may not be correct and that the stepwise regression results may be incorrect. The correlation between factors can be removed by performing PCA, since the components of the PCA are always uncorrelated. However, PCA also is not the perfect method, since the components are harder to interpret than the original factors.

3.3.3 Principal Component Analysis

To remove the multicollinearity, I perform PCA on the ten factors. I construct ten components that are linear combinations of the ten existing factors that maximize the variability of these linear combinations. I centre and scale the factors such that they are zero-centred and have unit variance before the analysis takes place. To reduce the number of factors and thus reduce the problem of overfitting, I select the factors with enough explanatory power and therefore have a high eigenvalue. My criterion for this is that the principal component should explain at least 5% of the variance, so it should have an eigenvalue larger than 0.5. This results in four principal components that explain around 86% of the total variance of the ten factors. A scree plot that contains the eigenvalues of all principal components can be found in Appendix F.

Table 3 shows the principal component (PC) coefficients. The first PC reflects the difference between the size and momentum factors and the other factors and explains 33% of the variance. This means that a third can be explained by one component that includes parts of all factors, so all factors are indeed important, as Liu et al. (2022) showed. The second PC mainly is the difference between the three longer momentum factors and DMU and explains 30% of the variance. This shows that almost another third of the variance can be explained by momentum factors, which is logical since there are four momentum factors. The third PC mostly measures the difference between the market and price factors and explains 15% of the variance. Lastly,

[5] The proof of multicollinearity is in Appendix E

the fourth PC reflects mainly the market factor and explains 8% of the variance, showing that the market factor is indeed important.

Table 3: Principal components

Factor	PC1	PC2	PC3	PC4
RMRF	0.127	0.033	-0.490	0.859
SMB	-0.359	0.255	0.187	0.189
DMU	0.249	0.380	0.096	0.054
PRC	0.370	-0.180	0.508	0.233
MAXPRC	0.380	-0.143	0.516	0.245
MOM2	-0.256	-0.441	0.021	0.105
MOM3	-0.271	-0.442	0.042	0.087
MOM4	-0.236	-0.438	0.095	0.104
VOL	0.400	-0.281	-0.296	-0.198
STDVOL	0.397	-0.275	-0.304	-0.200

3.4 Returns to be explained

I use the mean returns on 5×5 size-momentum portfolios as the returns to be explained. These are the means of the weekly returns of these 25 value-weight portfolios. I construct the portfolios by sorting the cryptocurrencies into quintiles based on their size and momentum, as in Fama and French (2012). Table 4 reports the mean returns of the 25 portfolios, plus the difference between the loser and winner portfolios, and the difference between the small-cap and large-cap portfolios. It again indicates the presence of the reversal effect since four out of the five returns in the last column are significantly positive. That means that the loser portfolios outperform the winner portfolios. The last column also shows that the reversal effect is stronger for the small-cap portfolio than for the large-cap portfolio. Furthermore, the table supports the results already found for the size factor, since three of the five returns in the last row are significantly positive. The size effect increases from the winner to the loser portfolio.

Table 4: Mean returns on the 5 × 5 size-momentum portfolios

Size quintiles	Prior returns quintiles					
	Up	2	3	4	Down	Down-Up
Big	0.027	0.018	0.011	-0.001	0.000	-0.035 (-3.158)***
2	-0.005	0.011	0.008	0.007	0.017	0.019 (2.219)**
3	-0.002	0.015	0.006	0.016	0.040	0.042 (4.214)***
4	-0.026	0.015	0.018	0.020	0.060	0.088 (10.530)***
Small	-0.022	0.023	0.041	0.055	0.120	0.142 (12.717)***
Small-Big	-0.050 (-4.148)***	0.004 (0.484)	0.028 (2.905)***	0.056 (5.898)***	0.123 (12.048)***	

Significance at the 1% or 5% level is denoted by *** or **, respectively.

4 Methodology

This section starts with an explanation of the forward and backward selection in Section 4.1. Secondly, Section 4.2 explains the models I investigate. Thirdly, the statistics used for model comparison are discussed in Section 4.3. Lastly, Section 4.4 explains the robustness checks.

4.1 Forward and backward selection

I perform forward and backwards selection to find the factors that have the most explanatory power. These two methods select the significant factors that increase the model's performance. I measure the performance by testing whether the intercepts of the regressions described in Section 4.2 are jointly equal to zero. If the model's factors completely explain the portfolio's average return, then the intercept is zero. That means the model performs well if the intercepts of the 25 regressions are jointly not significantly different from zero. I test this using the test statistic of the GRS test (further explanation of the GRS test is in Section 4.3). I choose the model with the lowest test statistic when comparing two models.

For forward selection, the procedure is as follows. I start with a regression that only contains the intercept. For every factor, I test the significance of that factor in 25 portfolios. If the factor is significant for at least half of the portfolios, it is a candidate to be added to the model. Then the GRS test statistic is computed for all the candidates. I add the candidate factor with the lowest test statistic to the model. The procedure repeats itself for the remaining factors, only this time, the new factor is already added to the model. The procedure stops when there are no significant factors left. This method gives the forward selection model.

The backward selection procedure is almost the same as the forward selection procedure, except that this method starts with a regression that contains all factors. This time the factors that are insignificant in at least half of the portfolios are candidates to be removed from the model. Again the GRS test statistic is computed for all candidates, and I remove the factor that decreases the test statistic the most when excluded. The procedure repeats itself for the remaining factors, only this time, the removed factor will be excluded from the model. The procedure stops when there are no insignificant factors left. This method gives the backward selection model.

It should be noted that the results from forward and backward selection are not very reliable, certainly not when there is multicollinearity. When performing forward and backward selection, the standard errors are biased toward zero and thus the p-values as well, as noted by Harrell (2010). This holds especially for the backward selection method. The backward selection starts with all factors included, so there is a lot of multicollinearity because of the high correlations[6]. That means that the p-values can be misleading and thus that irrelevant factors are selected. This also holds for the forward selection method when a candidate is highly correlated to a factor that is already included in the model, then the p-values can also be misleading. The Lasso explained in Section 4.4 can be an alternative to these two stepwise regression methods.

4.2 Models

There are in total five models that I test. I compare the five models to the C-CAPM, which is the cryptocurrency version of the CAPM. All models with the factors it includes are reported in Table 5.

Table 5: Model overview

Model	Factors
C-CAPM	RMRF
Three-factor	RMRF, SMB, DMU
Ten-factor	RMRF, SMB, DMU, PRC, MAXPRC, MOM2, MOM3, MOM4, VOL, STDVOL
PCA	Four linear combinations of the ten factors
Forward selection	The factors selected using forward selection
Backward selection	The factors selected using backward selection

[6] As proven in Appendix E

The general regression for the six models is as follows:

$$r_{it} - Rf_t = \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{f}_t + \varepsilon_{it}, \tag{5}$$

where r_{it} is a weekly return on one of the 5×5 size-momentum portfolios, Rf_t is the risk-free rate and $\boldsymbol{f_t}$ is a vector that contains the factors of the model that is tested. So, for each model, I perform 25 regressions that will be analyzed based on average and joint statistics.

4.3 Model performance

I compare the six models based on the average absolute intercept, average R^2, average adjusted R^2, average standard error of the intercepts and GRS p-value (as inspired by Shen et al. (2020)). The average absolute intercept indicates whether the model fully explains the cross-section of cryptocurrencies returns. A lower average absolute intercept indicates better performance. The R^2 shows the model's explanatory power, and the adjusted R^2 corrects for the difference in the number of factors. Lastly, the GRS p-value quantifies whether the intercepts are jointly significantly different from zero or not.

The GRS test is founded by Gibbons et al. (1989) and tests whether the intercepts in equation (5) are jointly equal to zero. That is:

$$H_0 : \alpha_i = 0, \forall i = 1, ..., N, \tag{6}$$

where α_i is the intercept and N is the number of portfolios. To compute the GRS test statistic, I define:

$$\hat{\Omega} = T^{-1} \sum_{t=1}^{T} \boldsymbol{f}_t \boldsymbol{f}_t', \tag{7}$$

and

$$\hat{\Sigma} = (T - L - 1)^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t', \tag{8}$$

where $\hat{\varepsilon}_t = (\hat{\varepsilon}_{1t}, ..., \hat{\varepsilon}_{Nt})'$ with $\hat{\varepsilon}_{it}$ as the OLS estimate of ε_{it} from equation (5), and L is the number of factors of the model. The equation of the GRS test statistic is as follows:

$$\frac{T(T - N - L)}{N(T - L - 1)} (1 + \bar{f}' \hat{\Omega}^{-1} \bar{f})^{-1} \hat{\boldsymbol{\alpha}}' \hat{\Sigma}^{-1} \hat{\boldsymbol{\alpha}}, \tag{9}$$

where $\hat{\boldsymbol{\alpha}}$ is the OLS estimate of $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)'$ and $\bar{f} = T^{-1} \sum_{t=1}^{T} \boldsymbol{f}_t$. The test statistic follows a $F(N, T - N - L)$ distribution under the H_0. The test assumes that $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ has a joint normal distribution with mean zero and nonsingular variance-covariance matrix Σ

and is i.i.d. over t. However, simulation evidence by MacKinlay (1986) suggests that the F test is fairly robust to a misspecification of the error terms.

4.4 Robustness checks

As a robustness check for the model performance, I look at the average absolute intercept, R^2 and the GRS p-value of the C-CAPM, three-factor and ten-factor model over time. This shows whether one model is always better than the other or that it is only in specific periods. Hence, I perform rolling regressions weekly with a window length of 52 weeks (as inspired by Shen et al. (2020)).

Furthermore, as a robustness check for the factor selection methods, I use the Lasso (least absolute shrinkage and selection operator) by Tibshirani (1996). Lasso is a variable selection and a regularization technique. Regularization is a technique that makes small changes to the learning algorithm to reduce overfitting. Lasso uses shrinkage, meaning that regression coefficients are shrunk toward a central point. The Lasso procedure can shrink coefficients of the variables with a minor contribution toward zero, such that it selects the relevant variables. It does so by penalizing the regression model with a penalty term called the L1-norm in the following regression:

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \tag{10}$$

where $\lambda \sum_{j=1}^{p} |\beta_j|$ is the L1-norm and λ denotes the amount of shrinkage. $\lambda = 0$ means that the penalty effect has no effect and that the regression is just the regular least squares regression. When λ increases, the coefficients will shrink toward zero. I choose λ by using ten-fold cross-validation. Cross-validation is a method that uses different parts of the data for testing and training, in different iterations. It trains on all but the k^{th} part, and then validates on the k^{th} part, where $k = 1, ..., 10$. This procedure then tries to find the best λ.

Lasso is useful for my research since there is a lot of multicollinearity, and regularization and factor reduction techniques, like Lasso, can help correct for this Schreiber-Gregory (2018). I use the Lasso on 5 of the 25 size-momentum portfolios, which are: big-up, big-down, 3-3, small-up and small-down.

5 Results

Section 5.1 first describes the results of the stepwise regression methods. Thereafter, Section 5.2 compares the six models that have been introduced. Lastly, Section 5.3 shows the results of the robustness checks.

5.1 Forward and backward selection

The forward and backward selection methods result in two different models. The forward selection method adds the following factors in the following order: MOM4, PRC, SMB, STDVOL, RMRF and DMU. So, this model includes the same factors as the three-factor model and adds MOM4, PRC and STDVOL. The backward selection model removes the following factors in the following order: MOM3, PRC, VOL, MOM2, SMB and MOM4. As a result, this model includes RMRF, DMU, MAXPRC and STDVOL. So, the backward selection model does not include the SMB factor, which is included in the three-factor model. It also finds a price factor and the standard deviation of the volume as factors that have significant explanatory power on the returns, like the forward selection method.

5.2 Model comparison

Table 6: Summary statistics for the regressions

	$\|\alpha\|$	R^2	Adj. R^2	s(α)	GRS p-value
C-CAPM	0.018	0.371	0.370	0.006	< 0.001
Three	0.020	0.417	0.412	0.008	< 0.001
Ten	0.017	0.456	0.443	0.009	< 0.001
PCA	0.018	0.426	0.421	0.009	< 0.001
Forward	0.016	0.446	0.438	0.009	< 0.001
Backward	0.020	0.438	0.432	0.008	< 0.001

The statistics of the regressions to compare the six models are in Table 6. It includes the average absolute intercepts, average R^2, average adjusted R^2, average standard error of the intercepts and GRS p-value. Appendix G reports the intercepts of all regressions with their t-statistic. Table 6 shows that all models outperform the C-CAPM in terms of the (adjusted) R^2. However, all models cannot completely explain the cross-section of cryptocurrency returns, since the p-values of the GRS test are significant on the 1% level. So, the models are incomplete descriptions of the cryptocurrency returns. The model that has the most explanatory power is the ten-factor model, since it has the highest (adjusted) R^2.

5.3 Robustness checks

Figure 1: Rolling graph of absolute intercepts

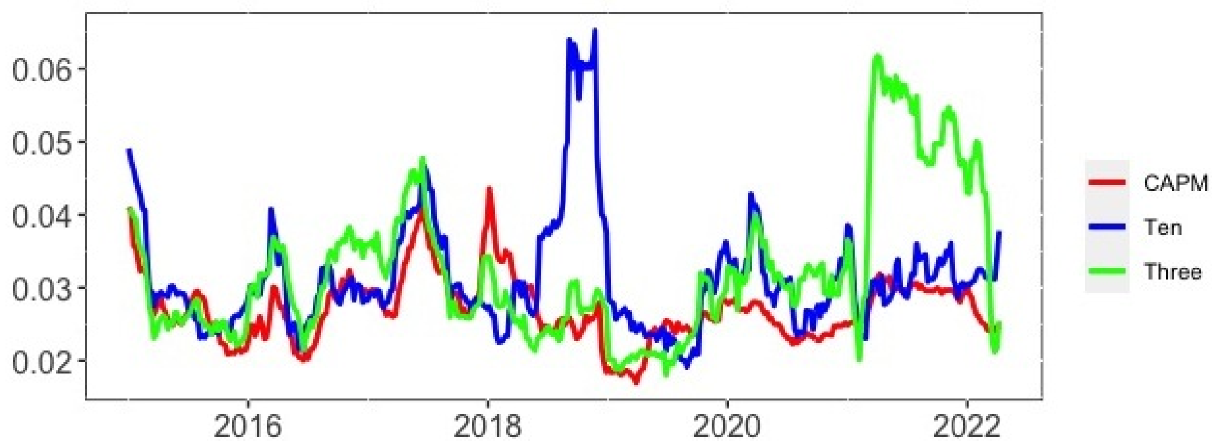


Figure 2: Rolling graph of R^2

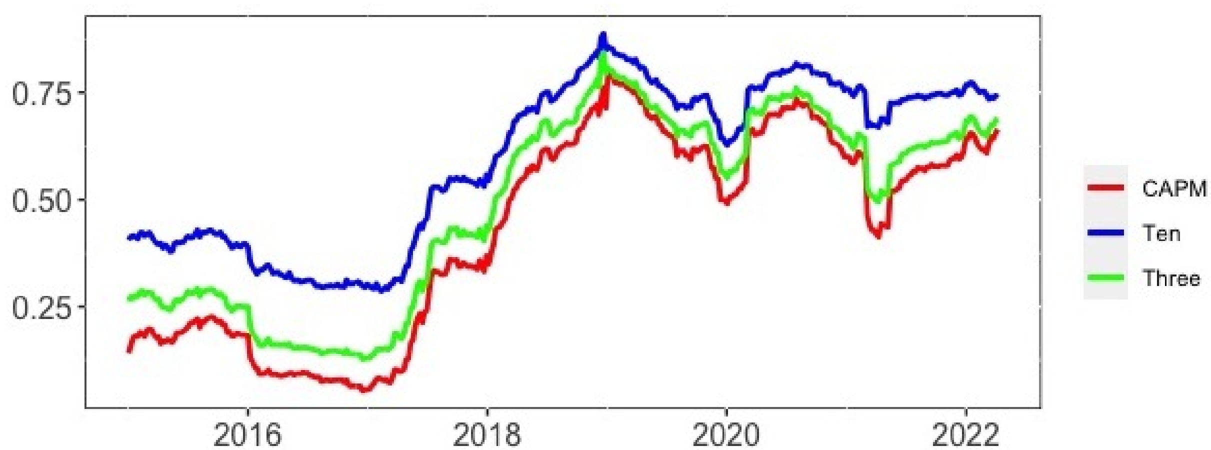
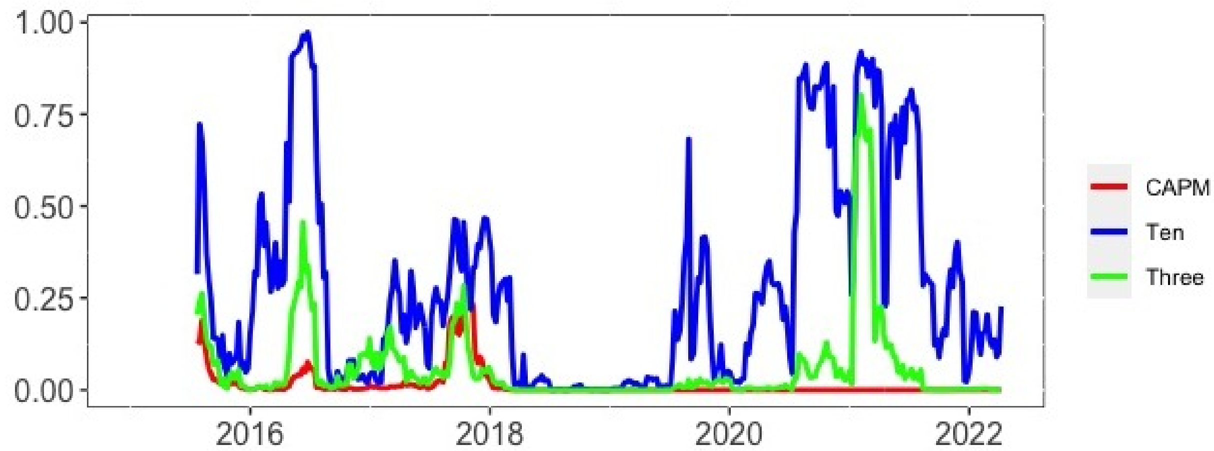


Figure 3: Rolling graph of GRS p-value



The results of the robustness check that uses a rolling window are in Figures 1, 2 and 3. Figure 1 shows that the average absolute intercept is not in all regressions lower for one model. For 77% of the periods, the average absolute intercept of CAPM is lower than that of the ten-factor model, especially around 2019 it is lower. Around 2022 is the average absolute intercept of the three-factor model higher than of the other two models. For the other periods, it is similar for all three models, so no model clearly outperforms the others in terms of the average absolute intercept. However, Figures 2 and 3 show that the ten-factor model is better. Its average R^2 is always larger than the average R^2 of C-CAPM and the three-factor model [7]. That means the ten-factor model's explanatory power is higher in all periods. Figure 3 shows the p-values of the GRS tests. In most periods, the p-value of the ten-factor model is much higher than the p-value of C-CAPM and the three-factor model. Furthermore, in 71% of the periods, it is higher than 0.05, meaning that the intercepts are not jointly significantly different from zero. Hence, the model can often explain the cross-section of cryptocurrency returns when looking at a window of one year. The three-factor model can explain the cross-section of cryptocurrency returns in 28% of the periods.

Table 7 reports the results of the Lasso on the five size-momentum portfolios. It shows that RMRF is selected for all five portfolios and has a high coefficient compared to the other factors.

MAXPRC, MOM2 and MOM3 are only selected for one of the five portfolios and have small coefficients, so they are irrelevant. MOM2 and MOM3 are also not selected by the stepwise methods, and MAXPRC is selected only by backward selection, so they find somewhat the same

[7]This result is the same for the adjusted R^2

results.

SMB has large positive coefficients for small-up and small-down, which is as expected, since it measures the returns of the small-caps minus the returns of the large-caps. However, SMB has a small negative coefficient for big-up, and for big-down it is not included in the model. So, for these two portfolios, SMB is not very relevant. Of the two stepwise regression methods, only forward selection finds SMB as a significant factor, so also not a very clear result for SMB.

DMU has large negative coefficients for the up portfolios and large positive returns for the down portfolios, again as expected, since DMU measures the returns of the losers minus the returns of the winners. Both stepwise methods also find DMU as a significant factor.

PRC has a large negative coefficient for big-down and small coefficients for big-down and small-up. The forward selection method also finds PRC as a significant factor contributing to the model.

MOM4 has three small coefficients, and one large positive coefficient for small-up, which shows that small-up has some longer momentum effects. The forward selection also finds MOM4 as a significant factor. So, MOM4 has some explanatory power, even when DMU is included.

VOL and STDVOL have relatively high positive coefficients for big-up and big-down. VOL also has a high negative coefficient for small-up. This shows that the volume factors are also relevant, as the stepwise selection methods also show.

Table 7: Lasso regression

Size quintile	Big	Big	3	Small	Small
Momentum quintile	Up	Down	3	Up	Down
RMRF	0.976	0.701	0.825	0.842	0.883
SMB	-0.051	-	-	0.278	0.176
DMU	-0.281	0.238	-	-0.244	0.500
PRC	-	0.093	-	-0.084	-0.241
MAXPRC	-	0.001	-	-	-
MOM2	-	-	-	-	-0.109
MOM3	-	-	-	0.138	-
MOM4	0.090	-	-0.004	0.209	-0.044
VOL	-	0.135	0.059	-0.252	-0.011
STDVOL	0.325	0.246	0.013	-0.086	-

6 Conclusion and discussion

This research tried to find a model that can better explain the cross-section of cryptocurrency returns and the factors contributing to this. For this, I used the three-factor model by Shen et al. (2020) as the base model. This model includes a market, size and reversal factor. First, I proved that the reversal effect is present in my data set and that it is the strongest for small-cap cryptocurrencies. It generates the highest return when the formation and holding period are both one week. I extended the three-factor model with seven factors proposed by Liu et al. (2022). This forms the ten-factor model, which outperforms the other models in terms of the (adjusted) R^2. I expected that it would have the highest R^2 since it includes the most factors, but corrected for the number of factors, it still has the most explanatory power. However, it cannot fully explain the cryptocurrency returns when considering the period from 2014 until now. Nonetheless, when looking at a window of one year, the ten-factor model is in most periods able to almost completely explain the cross-section of cryptocurrency returns. This is not what I expected, since the ten-factor model cannot explain the returns for the whole period.

I performed forward and backward selection on the ten factors to find the most relevant factors with relatively the most explanatory power. The forward selection model includes the same factors as the three-factor model, and also the four-week momentum, price and standard deviation of the volume factor. The backward selection method also selects a price factor and the standard deviation of the volume factor as significant factors, besides the market and reversal factor. As a result, both methods found that price and volume factors significantly contribute to the model. The Lasso procedure also found that the four-week momentum has some explanatory power when included together with the reversal factor. Furthermore, it also showed that the volume and price factors are relevant. I expected that the market size and reversal factor are important, as, for example, Shen et al. (2020) and Liu et al. (2020) showed. However, I did not expect that the four-week momentum would still add explanatory power and that the price and volume factors are still important.

To quantify the performance of the models, I used the GRS test. This test tests whether a set of factors can explain the returns of several portfolios. So, besides looking at the average absolute intercepts of the different models, like Shen et al. (2020) did, I analyzed the intercepts by testing whether they are jointly different from zero. This method can prove that a model can fully explain the cross-section of cryptocurrency returns, instead of indicating it.

The forward and backward selection methods that try to find the most important factors are not the best methods for this purpose. The resulting p-values can be misleading, so these two methods' conclusions are unreliable. This is especially the case with the high correlations and thus multicollinearity in my factors. That is why the Lasso procedure can help, because it deals better with multicollinearity than the stepwise selection methods and functions as a factor selection method. Furthermore, it reduces the problem of overfitting, such that the results are more applicable to other data sets. Since these three different methods conclude mostly the same things, the results are more reliable than if I only used one of these three methods.

The Principal Component Analysis (PCA) was less effective than I thought it would be. The PCA was useful for removing the multicollinearity, but the components are hard to interpret. This is especially a problem for one of the purposes that I used PCA for: factor selection. I thought there would be a more apparent distinction between the components so that I could identify important factors.

This research illustrates that the ten-factor model can explain the cryptocurrency returns in some periods, but that raises the question of the performance of the factors over time. It could be that, for example, the reversal effect changes over time or that its explanatory power is time-varying. Related to this is the question whether the predictability of the ten-factor model can be explained by, for example, the volatility of the returns. These two questions could be addressed in further research.

The research of this paper can be extended by including more cryptocurrencies. In total, there are over 13.000 cryptocurrencies available, so 11.000 more than that are included in my research. These mainly include coins with a small market capitalization, since they had their Initial Coin Offering (ICO) after 2019. These small currencies could change the results regarding the effects on the small-cap portfolios. Moreover, the returns after an ICO can also be investigated. It could be that there is a pattern in the returns in the period after an ICO.

Another interesting topic to use my research on, is looking at trading strategies using the factors that have a significant effect. This paper focuses only on the explanatory power of the factors. However, investigating whether these factors are useful for trading strategies is also relevant. The fact that factors perform well in-sample in explaining returns does not mean that they are also good at predicting returns. If the returns of cryptocurrencies can be predicted better, trading strategies that involve cryptocurrencies can be used by institutional investors.

References

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.

Chen, A. Y. and Zimmermann, T. (2020). Publication bias and the cross-section of stock returns. *The Review of Asset Pricing Studies*, 10(2):249–289.

Chuen, D. L. K., Guo, L., and Wang, Y. (2017). Cryptocurrency: A new investment opportunity? *The Journal of Alternative Investments*, 20(3):16–40.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 105(3):457–472.

Feng, G., Giglio, S., and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3):1327–1370.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, pages 1121–1152.

Grobys, K. and Sapkota, N. (2019). Cryptocurrencies and momentum. *Economics Letters*, 180:6–10.

Harrell, F. E. (2010). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer New York, 1 edition.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). . . . and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

Hou, K., Xue, C., and Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, 33(5):2019–2133.

Jain, P. K., McInish, T. H., and Miller, J. L. (2019). Insights from bitcoin trading. *Financial Management*, 48(4):1031–1048.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91.

Jia, Y., Liu, Y., and Yan, S. (2021). Higher moments, extreme returns, and cross–section of cryptocurrency returns. *Finance Research Letters*, 39:101536.

Lintner, J. (1965). The valuation of risk assets on the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(3):13–37.

Liu, W., Liang, X., and Cui, G. (2020). Common risk factors in the returns on cryptocurrencies. *Economic Modelling*, 86:299–305.

Liu, Y., Tsyvinski, A., and Wu, X. (2022). Common risk factors in cryptocurrency. *The Journal of Finance*, 77(2):1133–1177.

Lyandres, E., Palazzo, B., and Rabetti, D. (2019). Do tokens behave like securities? an anatomy of initial coin offerings. Working paper, SSRN Electronic Journal.

MacKinlay, A. C. (1986). An analysis of multivariate financial tests.

Menkhoff, L., Sarno, L., Schmeling, M., and Schrimpf, A. (2012). Currency momentum strategies. *Journal of Financial Economics*, 106(3):660–684.

Miffre, J. and Rallis, G. (2007). Momentum strategies in commodity futures markets. *Journal of Banking & Finance*, 31(6):1863–1886.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*, page 21260.

Petukhina, A., Trimborn, S., Härdle, W. K., and Elendner, H. (2020). Investing with cryptocurrencies–evaluating their potential for portfolio allocation strategies. *Quantitative Finance*, 2021:1–29.

Rouwenhorst, K. G. (1998). International momentum strategies. *The Journal of Finance*, 53(1):267–284.

Schreiber-Gregory, D. N. (2018). Ridge regression and multicollinearity: An in-depth review. *Model Assisted Statistics and Applications*, 13(4):359–365.

Schwenkler, G. and Zheng, H. (2020). Peer co-movement in crypto markets. *Available at SSRN 3572471*.

Shahzad, S. J. H., Bouri, E., Ahmad, T., Naeem, M. A., and Vo, X. V. (2021). The pricing of bad contagion in cryptocurrencies: a four-factor pricing model. *Finance Research Letters*, 41:101797.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Shen, D., Urquhart, A., and Wang, P. (2020). A three-factor pricing model for cryptocurrencies. *Finance Research Letters*, 34:101248.

Steer, G. (2022). Did institutional investors crash the crypto party? *Financial Times*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tzouvanas, P., Kizys, R., and Tsend-Ayush, B. (2020). Momentum trading in cryptocurrencies: Short-term returns and diversification benefits. *Economics Letters*, 191:108728.

Zhang, W. and Li, Y. (2020). Is idiosyncratic volatility priced in cryptocurrency markets? *Research in International Business and Finance*, 54:101252.

Appendix A All 25 factors

- Market capitalization
- Price
- Maximum price
- Number of weeks listen
- One-week momentum
- Two-week momentum
- Three-week momentum
- Four-week momentum
- Eight-week momentum
- Sixteen-week momentum
- Fifty-week momentum
- Hundred-week momentum
- Volume
- Volume times price
- Volume times price scaled by market capitalization
- The regression coefficient β_i in $r_i - Rf = \alpha_i + \beta_i MKT + \varepsilon_i$
- Beta squared
- Idiosyncratic volatility
- Standard deviation of daily returns
- Skewness of daily returns
- Kurtosis of daily returns
- Maximum daily return
- Improvement of R^2 in $r_i - Rf = \alpha_i + \beta_{i,MKT} MKT + \beta_{i,MKT_{-1}} MKT_{-1} + \beta_{i,MKT_{-2}} MKT_{-2} + \varepsilon_i$, compared to using only current coin market excess returns
- Standard deviation of volume
- Average absolute daily return divided by volume

Appendix B Stablecoins

- Thether
- USD Coin
- TrueUSD

- Pax Dollar
- Gemini Dollar
- Stasis Euro
- sUSD
- BitCNY
- Stably USD

Appendix C Mean returns J-K portfolios

Table 8: Mean returns on the J-K portfolios

J		K = 1	2	3	4
1	Buy	-0.033***	-0.036***	-0.043***	-0.052***
		(-5.150)	(-4.155)	(-3.998)	(-4.299)
1	Sell	0.121***	0.106***	0.101***	0.088***
		(17.194)	(11.747)	(9.185)	(7.287)
1	Buy-Sell	-0.154***	-0.143***	-0.143***	-0.140***
		(-21.373)	(-16.763)	(-16.029)	(-15.571)
2	Buy	-0.021***	-0.023**	-0.025**	-0.036***
		(-3.091)	(-2.456)	(-2.283)	(-2.897)
2	Sell	0.109***	0.111***	0.101***	0.094***
		(14.925)	(12.105)	(9.218)	(7.700)
2	Buy-Sell	-0.130***	-0.133***	-0.126***	-0.130***
		(-16.549)	(-14.924)	(-14.724)	(-13.498)
3	Buy	-0.012*	-0.013	-0.023**	-0.035***
		(-1.872)	(-1.494)	(-2.129)	(-2.901)
3	Sell	0.099***	0.100***	0.092***	0.079***
		(13.797)	(10.971)	(8.292)	(6.856)
3	Buy-Sell	-0.111***	-0.113***	-0.114***	-0.114***
		(-15.797)	(-14.253)	(-13.630)	(-13.125)
4	Buy	-0.008	-0.011	-0.015	-0.023*
		(-1.206)	(-1.219)	(-1.315)	(-1.776)
4	Sell	0.100***	0.102***	0.092***	0.082***
		(14.464)	(11.190)	(8.362)	(6.946)
4	Buy-Sell	-0.107***	-0.133***	-0.107***	-0.104***
		(-15.557)	(-14.557)	(-12.477)	(-10.688)

The *t*-statistics are reported in parentheses. Significance at the 1%, 5% or 10% level is denoted by ***, ** or *, respectively.

Appendix D Descriptive statistics and correlations of factors

Table 9: Descriptive statistics

	Mean	St. Dev.	Skewness	Kurtsosis
RMRF	0.000	0.104	0.058	4.736
SMB	0.063	0.137	0.322	5.456
DMU	0.061	0.083	0.751	13.718
PRC	-0.028	0.086	-0.667	6.709
MAXPRC	-0.027	0.086	-0.793	7.059
MOM2	-0.089	0.112	-0.864	15.464
MOM3	-0.081	0.102	0.925	5.705
MOM4	-0.082	0.095	0.843	5.883
VOL	-0.027	0.102	-0.235	5.717
STDVOL	-0.029	0.103	-0.269	5.985

Table 10: Correlations

	RMRF	SMB	DMU	PRC	MAXPRC	MOM2	MOM3	MOM4	VOL	STDVOL
RMRF	1.000									
SMB	-0.137	1.000								
DMU	0.088	0.066	1.000							
PRC	-0.071	-0.419	0.007	1.000						
MAXPRC	-0.061	-0.404	0.051	0.970	1.000					
MOM2	-0.090	-0.072	-0.731	0.038	0.001	1.000				
MOM3	-0.133	0.013	-0.612	-0.055	-0.104	0.763	1.000			
MOM4	-0.144	-0.031	-0.537	0.011	-0.035	0.711	0.819	1.000		
VOL	0.208	-0.687	-0.038	0.392	0.366	-0.007	0.003	0.006	1.000	
STDVOL	0.211	-0.679	-0.044	0.387	0.363	0.002	0.000	0.011	0.959	1.000

Appendix E Proof of multicollinearity

Multicollinearity can be measured using the Variance Inflation Factor (VIF). The VIF estimates how much the variance of a regression coefficient is inflated due to multicollinearity. A large

VIF on a factor indicates high multicollinearity. The VIF is computed as follows:

$$VIF_j = \frac{1}{1 - R_j^2}, \tag{11}$$

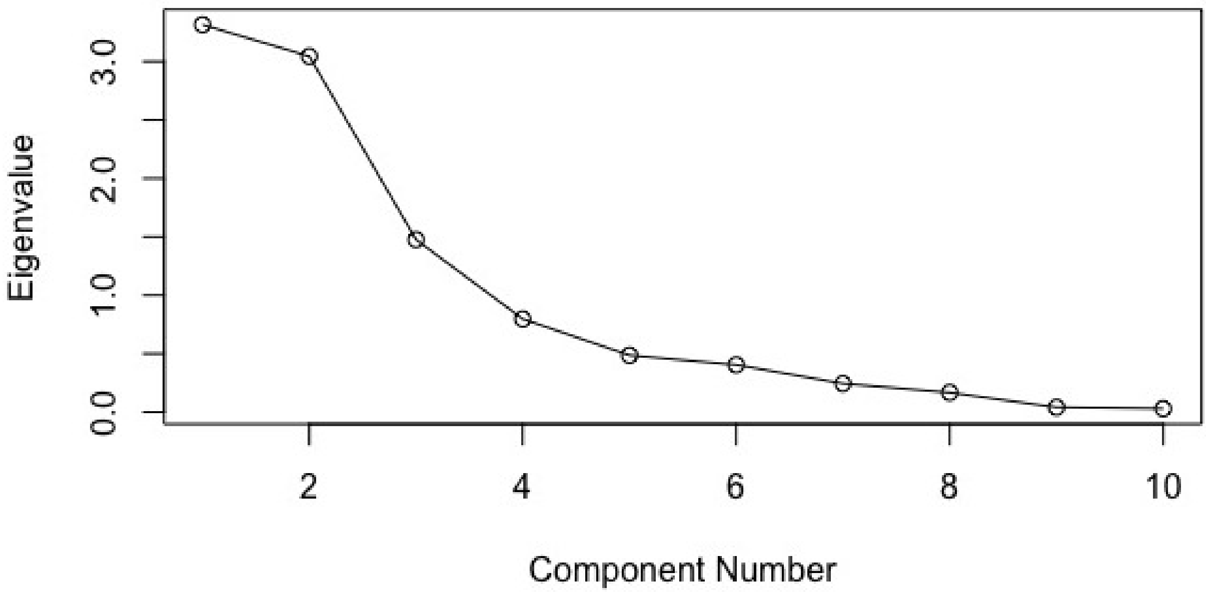
where R_j^2 is the R^2 obtained by regressing the j^{th} factor on the other factors. The VIF values for the regression with the big-up portfolio as the dependent variable are below in Table 11. A VIF of one means no multicollinearity and a VIF higher than 10 indicates serious multicollinearity. The table shows that there is almost no multicollinearity in the three-factor, forward selection and backward selection models. However, there is serious multicollinearity in the ten-factor model. The multicollinearity is not a very big problem for the momentum factors, despite the high correlations.

Table 11: VIF values

Factor	Three-factor	Ten-factor	Forward	Backward
RMRF	1.022	1.094	1.090	1.072
SMB	1.027	2.010	1.929	
DMU	1.011	1.981	1.421	1.014
PRC		18.727	1.294	
MAXPRC		18.462		1.185
MOM2		3.011		
MOM3		4.120		
MOM4		3.219	1.401	
VOL		13.852		
STDVOL		15.529	1.978	1.232

Appendix F PCA

Figure 4: Scree plot



Appendix G Intercepts regressions

The tables below show the 25 intercepts of the six models and their t-statistics.

Table 12: Intercepts from C-CAPM

	α					$t(\alpha)$				
	Up	2	3	4	Down	Up	2	3	4	Down
Big	0.014	0.005	-0.003	-0.014	-0.015	1.812	1.149	-0.724	-3.335	-1.895
2	-0.018	-0.002	-0.005	-0.007	0.000	-2.581	-0.389	-1.123	-1.572	0.059
3	-0.016	0.001	-0.008	0.003	0.026	-1.890	0.208	-1.489	0.538	3.868
4	-0.040	0.001	0.004	0.007	0.046	-5.811	0.189	0.658	1.177	7.356
Small	-0.036	0.009	0.027	0.040	0.106	-3.967	1.281	2.899	4.836	14.079

Table 13: Intercepts from three-factor model

	α					$t(\alpha)$				
	Up	2	3	4	Down	Up	2	3	4	Down
Big	0.052	0.015	-0.003	-0.023	-0.051	5.206	2.556	-0.671	-4.044	-4.846
2	0.025	0.013	-0.007	-0.008	-0.027	2.792	1.977	-1.181	-1.296	-3.723
3	0.028	0.017	-0.005	-0.008	-0.002	2.728	2.274	-0.792	-1.172	-0.283
4	-0.023	0.018	-0.006	-0.009	0.013	-2.599	2.325	-0.673	-1.256	1.633
Small	-0.039	0.006	0.026	0.014	0.052	-3.609	0.624	2.098	1.270	5.830

Table 14: Intercepts from ten-factor model

	α					$t(\alpha)$				
	Up	2	3	4	Down	Up	2	3	4	Down
Big	0.061	0.008	-0.008	-0.023	-0.020	5.307	1.254	-1.376	-3.426	-1.759
2	0.026	0.007	-0.017	-0.012	-0.033	2.755	0.953	-2.343	-1.792	-4.229
3	0.018	0.006	-0.006	-0.012	0.004	1.578	0.687	-0.785	-1.612	0.411
4	-0.024	0.009	-0.014	-0.014	0.001	-2.365	1.061	-1.471	-1.741	0.139
Small	-0.019	-0.006	0.023	0.019	0.043	-1.490	-0.516	1.698	1.523	4.111

Table 15: Intercepts from forward selection model

	α					$t(\alpha)$				
	Up	2	3	4	Down	Up	2	3	4	Down
Big	0.061	0.012	-0.005	-0.024	-0.019	5.377	1.802	-0.995	-3.657	-1.798
2	0.024	0.009	-0.015	-0.011	-0.031	2.572	1.228	-2.195	-1.658	-4.028
3	0.015	0.008	-0.007	-0.012	0.001	1.279	0.962	-0.973	-1.579	0.136
4	-0.025	0.008	-0.012	-0.011	0.002	-2.496	0.889	-1.292	-1.342	0.239
Small	-0.019	-0.003	0.020	0.017	0.040	-1.581	-0.306	1.497	1.404	3.967

Table 16: Intercepts from backward selection model

	α					$t(\alpha)$				
	Up	2	3	4	Down	Up	2	3	4	Down
Big	0.047	0.012	-0.011	-0.023	-0.054	4.809	2.037	-2.396	-4.103	-5.317
2	0.028	0.013	-0.010	-0.006	-0.027	3.241	2.156	-1.633	-0.938	-3.759
3	-0.025	0.014	-0.009	-0.013	0.001	2.552	1.875	-1.337	-2.041	0.132
4	-0.025	0.011	-0.01	-0.014	0.004	-2.907	1.408	-1.293	-1.959	0.593
Small	-0.031	0.004	0.025	0.018	0.055	-2.859	0.396	1.986	1.717	6.116

Table 17: Intercepts from PCA model

	α					$t(\alpha)$				
	Up	2	3	4	Down	Up	2	3	4	Down
Big	0.054	0.007	-0.007	-0.020	-0.033	4.864	1.125	-1.358	-3.179	-2.668
2	0.029	0.014	-0.014	-0.005	-0.024	2.997	2.097	-2.118	-0.845	-3.025
3	0.025	0.007	-0.009	-0.007	0.010	2.236	0.851	-1.271	-0.928	1.086
4	-0.023	0.005	-0.011	-0.015	0.007	-2.333	0.587	-1.245	-1.913	0.815
Small	-0.021	-0.010	0.020	0.019	0.045	-1.814	-0.952	1.483	1.622	4.536

Appendix H Script description

- The "getData.R" document retrieves the data from the internet
- The "setData.R" document sets the data such that it can be used in the programming and sets some parameters. It also makes a table of the descriptive statistics of the data
- The "J-K portfolios.R" document computes everything for the process of determining the best J-K strategy
- The "factorModels.R" document computes all the factors, does the 5×5 portfolio calculations, does the forward and backward selection methods, computes all the necessary regressions and performs the robustness checks
- The "workspace.RData" document is the workspace that can be loaded into R that contains the data of the cryptocurrencies