Team 37 Final Report:
Factor Modeling of Cryptocurrency Returns

Team 37: Cameron Corbridge, Charles Kramer, Jianzhen (Jason) Liu, Tyler Meyers, and Neelam Saini

Can cryptocurrencies add value to portfolios of more traditional assets? To address this question, we examine returns on cryptocurrency portfolios for evidence of systematic risk factors associated with stock market returns, as well as for excess returns (Jensen's alpha). Our evidence is consistent with the idea that cryptocurrency portfolios are 'high beta' assets that offer excess returns. That said, caution is advised given the large number of outliers and missing values present in the data, and further research would be warranted to determine whether cryptocurrency would be a suitable investment for risk-averse investors.

Background

Cryptocurrency emerged as a new form of digital currency in 2009 with the launch of Bitcoin. Since then, the cryptocurrency landscape has evolved rapidly with the introduction of numerous other digital currencies (aka 'coins') such as Ethereum, Ripple, and Litecoin. Cryptocurrency has introduced a new form of investment that allows for a combination of decentralization, anonymity, and security that is not available in traditional investments. Cryptocurrency has also facilitated cross-border transactions, making it easier and faster to transfer money across different countries.[6]

However, crypto's reputation as a purely speculative market, beset by illiquidity and volatility, has led many traditional stock-market investors to avoid it. While the overall market is impressive in size, it remains fragmented, with many small coins, and the institutional investment and the availability of investment instruments such as mutual funds are still limited [11,13]. In [7], the authors analyzed a dataset of over 800 cryptocurrency offerings and found that over 10% of them exhibited clear Ponzi-like characteristics. Skeptics point to the risks related to crypto's complexity and opacity as a drawback to potential investors[10].

But to the extent that the crypto market is neglected, this may create opportunities, much as the neglected market for 'fallen angel' corporate bonds led to the development of the high-yield bond market in the 1970s. Indeed, the emergence of cryptocurrency mutual funds, futures and options, while still nascent, suggests a place for cryptocurrency in the investors' toolkit. Thus we ask, does crypto reflect similar risks as investments such as stocks, and can they offer diversification or yield to a standard equity portfolio? The answer would be of great value for asset managers, institutional investors, and intermediaries that design new investment products.

Targets and Approach

To investigate whether cryptocurrencies add value to a portfolio of traditional securities, we will use standard portfolio analysis tools. We will measure underlying systematic risks via factors from Fama and French [14] and examine the extent to which crypto portfolio returns are explained by the factors. We use the formula

$$r_{it} - r_{ft} = \alpha_i + \sum_j \beta_{ij} f_{jt} + \epsilon_{it}$$

where $r_{it} - r_{ft}$ is the excess return (over the risk-free rate) of cryptocurrency portfolio i, $f_j = \{SMB, HML, Market{-}Risk\text{-}free\}$ are the returns on the Fama-French small-minus-big, high-minus-low book/market, and market excess return factors, and ϵ_{it} is the idiosyncratic risk in cryptocurrency portfolio i. The research question then comes down to

$$H_0\colon \beta_{ij} = 0 \text{ for all } j,$$

e.g., a joint test of significance of the factor responses. Intuitively, this tells us if cryptocurrencies contain the same risk factors as more common and familiar assets. For example, SMB could correlate with the relative illiquidity and volatility in cryptocurrency, given the lower liquidity and higher volatility of small versus large firms. HML could pick up deviations from fundamentals, given the propensity of cryptocurrencies to undertake large swings in value. Given the limited impact of macro news on cryptocurrencies [12], we expect that the coefficient on Market −Risk-free may be insignificant.

We use a one-sided test for a significant Jensen's alpha (excess return):

$$H_0 : \alpha_i = 0 \text{ vs } H_1 : \alpha_i > 0.$$

We perform a one-sided test because if alpha is negative, then crypto has negative value added. In principle, one could exploit this by shorting crypto, but not every cryptocurrency has the features such as derivatives or margin to facilitate short selling. In addition, short selling poses risks and administrative complexities, particularly for illiquid securities.

Additionally, we fit regressions to explore the impact of other variables, such as market volatility (VIX—implied volatility of the S&P 500 index[1]), Google Trends mentions of Bitcoin, and Twitter tweet volume on Bitcoin, on cryptocurrency returns.[2] By conducting these additional analyses, we hope to provide a more comprehensive understanding of the factors driving cryptocurrency returns and to identify potential sources of risk and return in this emerging asset class. We use the formula

$$r_{it} - r_{ft} = \alpha_i + \sum_j \beta_{ij} f_{jt} + \sum_k \gamma_{ik} g_{kt} + \varepsilon_{it},$$

where g_{kt} are these additional factors that we might find in crypto returns but not necessarily stock returns. Accordingly, we test

$$H_0 : \gamma_k = 0 \text{ vs } H_1 : \gamma_k \neq 0.$$

We evaluate our models using metrics such as R-squared and perform a battery of robustness tests (Variance Inflation Factor, serial correlation, heteroskedasticity, and structural break (Chow) tests). Based on our analysis, we expect to find evidence in favor of the hypothesis that cryptocurrencies add value to a portfolio of traditional securities, providing additional risk-adjusted return and better diversification. This could impact asset allocation decisions by investors, potentially leading to increased adoption and investment in cryptocurrencies.

[1] VIX is the volatility implied by options on the S&P500, and is commonly employed as a measure of variation in overall market risk.
[2] The data are available for individual cryptocurrencies; we use Bitcoin as a bellwether of the market, given its large market capitalization and broad name recognition.

Data

We cleaned and merged six datasets, which include:
Cryptocurrency Data

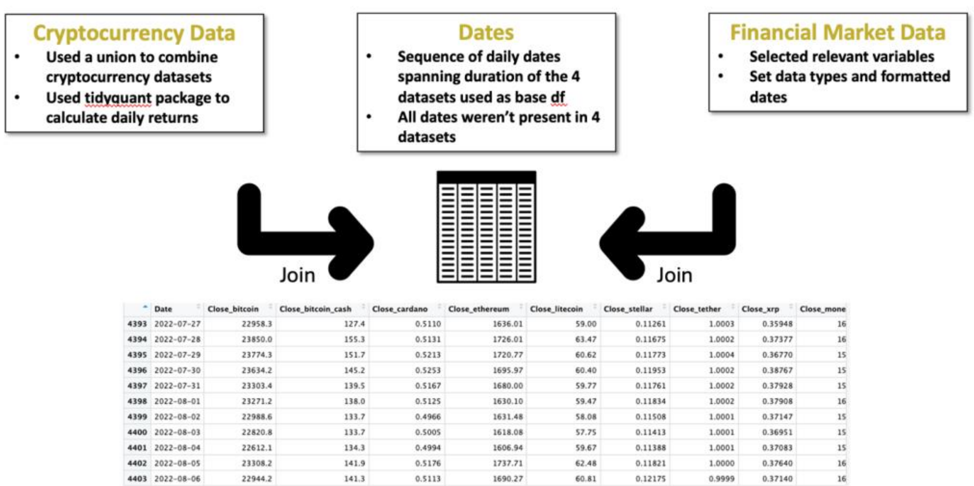
- Top 100 Cryptocurrencies Historical Dataset – Kaggle [1]
    - Selected top 30 cryptocurrencies by market cap
    - Variables – Close, Volume, calculated Daily Returns
- Complete Historical Cryptocurrency Financial Data – Kaggle [2]
    - Supplemental historical data for subset of top 30 cryptocurrencies
    - Variables – Close, Volume, calculated Daily Returns

Financial Markets Data

- VIX (CBOE Volatility Index) – Federal Reserve Economic Data [3]
    - Measure of market expectation of near-term volatility
    - Variables – VIXCLS
- Fama-French Research Factors - Ken French Data Library [4]
    - Measures of sensitivity to market, sensitivity to size, and sensitivity to value stocks
    - Variables – Mkt_Rf, SMB, HML, RF

Other Factors:

- Google Trends data on Bitcoin [15]
- Twitter mentions of Bitcoin [16]



	Date	Close_bitcoin	Close_bitcoin_cash	Close_cardano	Close_ethereum	Close_litecoin	Close_stellar	Close_tether	Close_xrp	Close_mone
4393	2022-07-27	22958.3	127.4	0.5110	1636.01	59.00	0.11261	1.0003	0.35948	16
4394	2022-07-28	23850.0	155.3	0.5131	1726.01	63.47	0.11675	1.0002	0.37377	16
4395	2022-07-29	23774.3	151.7	0.5213	1720.77	60.62	0.11773	1.0004	0.36770	15
4396	2022-07-30	23634.2	145.2	0.5253	1695.97	60.40	0.11953	1.0002	0.38767	15
4397	2022-07-31	23303.4	139.5	0.5167	1680.00	59.77	0.11761	1.0002	0.37928	15
4398	2022-08-01	23271.2	138.0	0.5125	1630.10	59.47	0.11834	1.0002	0.37908	16
4399	2022-08-02	22988.6	133.7	0.4966	1631.48	58.08	0.11508	1.0001	0.37147	15
4400	2022-08-03	22820.8	133.7	0.5005	1618.08	57.75	0.11413	1.0001	0.36951	15
4401	2022-08-04	22612.1	134.3	0.4994	1606.94	59.67	0.11388	1.0001	0.37083	15
4402	2022-08-05	23308.2	141.9	0.5176	1737.71	62.48	0.11821	1.0000	0.37640	16
4403	2022-08-06	22944.2	141.3	0.5113	1690.27	60.81	0.12175	0.9999	0.37140	16

Figure 1: Data Preparation Flow Chart

Our merged dataset is 4,420 observations in length, although there are many missing values for cryptocurrency prices (reflecting days when the currency did not trade). This affects portfolio formation, where we need returns for all the components every day when the portfolio return is measured. As a result, the portfolio records are much shorter in length.

There are more than 11,000 cryptocurrencies on the market, but many do not last long. The Cambridge Centre for Alternative Finance found that the median lifespan of a cryptocurrency is only 1.2 years [5]. Our analysis is accordingly subject to survivorship bias because we are starting with a sample of currently traded coins (implicitly excluding failed coins). This can bias our returns upward.

Selecting a subset of cryptocurrencies for studying risk and return measures is the first step we must take. We focus on the largest and most actively traded ones, which are likely to be the most appealing to investors. We exclude stablecoins, which are cryptocurrencies tied to the value of a currency or financial asset and focus only on price returns. We ignore other types of returns, such as proof of stake returns or returns on lending crypto.

We have selected the top 10 cryptocurrencies by market capitalization for analysis, excluding those that are stablecoins. This yields the following coins: Bitcoin (BTC), Ethereum (ETH), BNB (BNB), XRP (XRP), Cardano (ADA), Dogecoin (DOGE), Polygon (MATIC), Solana (SOL), Polkadot (DOT), and Litecoin (LTC).

Exploratory Data Analysis

We start our exploratory analysis with boxplots of price returns for each currency (Figure 1).

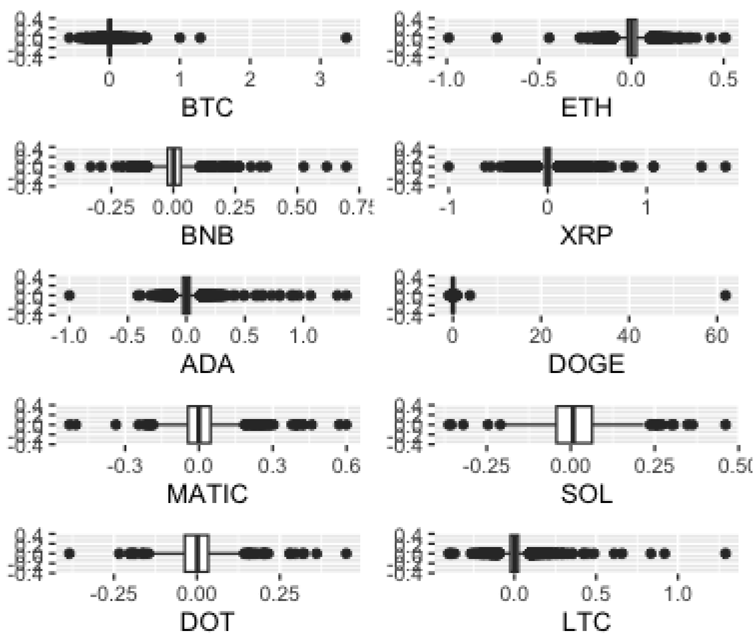


Figure 1: Box Plots of Crypto Returns

The box plots clearly show the presence of outliers—indeed for several currencies, the interquartile range is barely visible. However, upon examination, these outliers turn out to be bona fide returns. Excluding them could improve our results but at the cost of realism. Accordingly, we leave them in the data for our initial analysis, although we check later for their influence on the regressions.

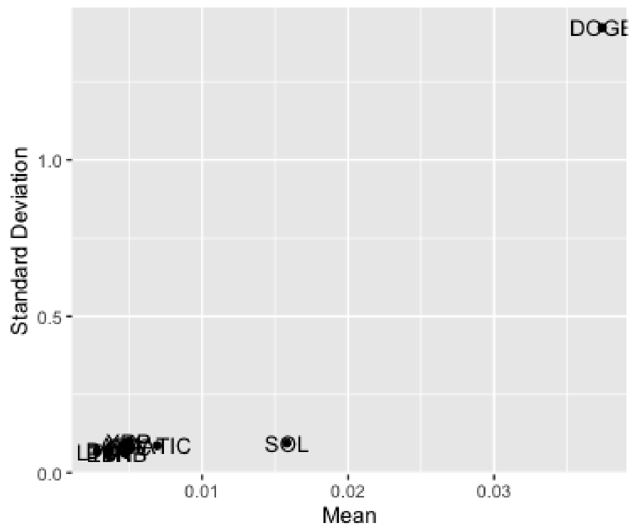


Figure 2: Mean and Standard Deviation of Crypto Returns

Figure 2 plots the mean and standard deviation for the selected coins. Consistent with the large outliers for DOGE, its mean and standard deviation are well outside those for the rest of the coins, which are clustered closer to the origin.

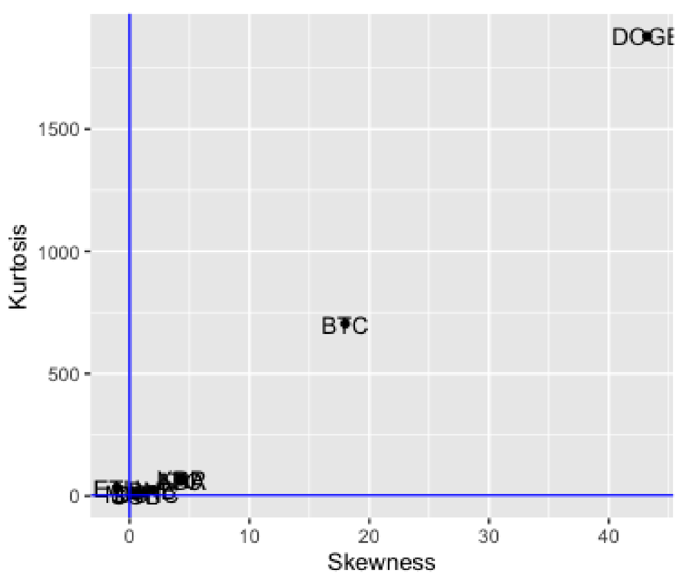


Figure 3: Skewness and Kurtosis (blue lines=values for normal distribution)

Similarly, skewness and kurtosis (Figure 3) are huge for DOGE, and quite large for BTC. Even for many of the other currencies, the excess kurtosis relative to the value for the normal distribution shows 'heavy-tailed' data that are the hallmark of stock returns.

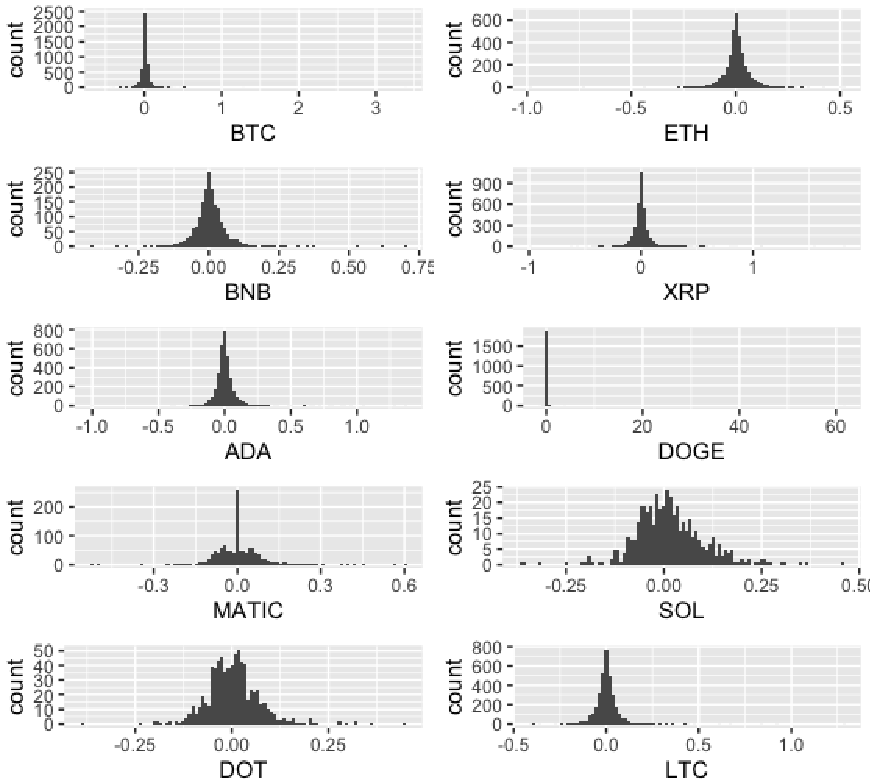


Figure 4: Distributions of Returns

The presence of outliers is also evident in the distribution plots (Figure 4), with most plots showing a highly peaked distribution and large outliers (most pronounced for DOGE).

Portfolio Analysis

We initiate our analysis with an examination of Sharpe ratios for our selected currencies, comparing them with the ratio for the market return.

Asset	Sharpe ratio
BTC	0.0538
ETH	0.0186
BNB	0.0230
XRP	0.0399
ADA	0.0105
DOGE	0.0340
MATIC	0.0842
SOL	0.1349
DOT	0.0306
LTC	-0.0031
Mkt	0.0499

Figure 6: Sharpe Ratios

Figure 6 shows the Sharpe ratio for our selected currencies along with the market return (Mkt). Note that the Sharpe ratios for a few cryptocurrencies are similar to or larger than the ratio for the market return (e.g. for BTC, MATIC, SOL), although many are smaller and LTC has a negative Sharpe ratio (e.g. negative mean excess return).

We next form three portfolios: equal-weighted, market-cap weighted, and principal components. The equal-weighted portfolio o assigns an equal weight of 1/10 to each cryptocurrency. Market-cap weights are based on market capitalization, while principal component weights are calculated using the *prcomp* function in R (we choose the first principal component).

We calculate daily portfolio returns by multiplying the weights with the daily returns of each cryptocurrency, then fit linear regression models to each portfolio. Table 1 shows our results for the basic model with factors Mkt-RF, SMB, HML.

Each of the portfolios has a significant coefficient on market risk (Mkt_RF), and each is greater than 1.0, implying that crypto portfolios are ‘high-beta’. This is particularly true for the principal components regression, with a beta of 5.4. The significant loading on market risk is surprising given the aforementioned research finding that cryptocurrencies do not have macro risks. That said, it is consistent with the understanding that cryptocurrencies are high-volatility assets. F-tests for joint significance of the slope coefficients all reject the null that the coefficients are jointly zero, likely a reflection of the large and significant loading on market risk.

It is also encouraging that both equal-weighted and principal-component portfolios have significant intercepts (Jensen’s alpha). That said, for the principal component portfolio the significance is low (10 percent on a two-tailed basis, or 5 percent on a one-tailed basis) and the magnitude is implausibly large (2.9 percent *per day*).

The remainder of the results are less strong. None of the coefficients on size (SMB) or value (HML) are significant, except for the principal component portfolio, which has a significant coefficient on SMB. However, the sign for this coefficient is the opposite of our hypothesized sign—e.g. rather than having a positive loading (indicative of shared characteristics with small companies), its loading is negative. R-squared is small—none of the regressions explain more than 10 percent of the variation in crypto portfolio returns.

Table 1. Factor Regressions for Crypto Portfolios

	Dependent variable:		
	Equal-Weighted	Market-Weighted	Principal Component
	(1)	(2)	(3)
Mkt_RF	1.890***	1.282***	5.371***
	(0.438)	(0.313)	(1.838)
SMB	0.082	0.570	-4.012*
	(0.499)	(0.357)	(2.094)
HML	-0.188	-0.244	0.341
	(0.320)	(0.229)	(1.344)
Constant	0.009**	0.005	0.029*
	(0.004)	(0.003)	(0.017)
Observations	256	256	256
R^2	0.084	0.102	0.038
Adjusted R^2	0.073	0.092	0.026
Residual Std. Error (df = 252)	0.064	0.046	0.269
F Statistic (df = 3; 252)	7.736***	9.579***	3.309**
Note: *p<0.1; **p<0.05; ***p<0.01			

We next added VIX as a factor (Table 2). No portfolio has a significant coefficient on VIX, and the coefficients on other factors are broadly unchanged, as are the test statistics for other factors, R-squared and adjusted R-squared. However, Jensen's alpha is no longer significant for any portfolio. We suspect that this result is spurious (as the mean of VIX may absorb the excess return in the portfolios). Indeed, removing the mean of VIX restores the significance of alpha for the equal-weighted portfolio (not shown).

We proceeded to perform robustness tests on the regressions (Table 3). Multicollinearity, heteroskedasticity and serial correlation can bias regression standard errors, reducing the accuracy of standard hypothesis tests. Variance inflation factors (row 1) for the factors are close to 1, indicating that multicollinearity is not an issue. Similarly, none of the three regressions show serial correlation in the residuals (row 2). However, there is significant heteroskedasticity in both the equal-weighted and principal-component regressions (row 3). In addition, the principal-component regression shows parameter instability (row 4).

The presence of significant outliers, as described above, may influence the results. Accordingly, for each of the three models, we calculated Cook's Distance for each observation and excluded those that were more than three times the mean Cook's Distance. Re-running the regressions results in only moderate changes to the model results, however (see Annex Table 1); SMB is no longer significant for the principal-components model, but other results are similar. The models purged of outliers no longer show parameter instability for the principal-components model, although heteroskedasticity remains an issue (and the equal-weighted portfolio model shows moderate evidence of serial correlation; see Annex Table 2).

Table 2. Factor Regressions with VIX

	Dependent variable:		
	Equal-Weighted	Market-Weighted	Principal Component
	(1)	(2)	(3)
Mkt_RF	1.863^{***}	1.290^{***}	5.874^{***}
	(0.451)	(0.323)	(1.891)
SMB	0.099	0.565	-4.327^{**}
	(0.504)	(0.361)	(2.112)
HML	-0.186	-0.244	0.316
	(0.321)	(0.230)	(1.344)
VIXCLS	-0.0002	0.0001	0.004
	(0.001)	(0.001)	(0.004)
Constant	0.014	0.003	-0.061
	(0.019)	(0.014)	(0.081)
Observations	256	256	256
R^2	0.085	0.102	0.043
Adjusted R^2	0.070	0.088	0.028
Residual Std. Error (df = 251)	0.064	0.046	0.269
F Statistic (df = 4; 251)	5.797^{***}	7.158^{***}	2.803^{**}
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 3. Diagnostic Statistics for Factor Regressions

1. VIF	**Mkt_RF** 1.2	**SMB** 1.2	**HML** 1.1	**VIX** 1.1
Robustness tests:				
	Equal-Weighted	**Market-Weighted**	**Principal Component**	
2. Serial Correlation	p = 0.232	p = 0.19	p = 0.134	
3. Heteroskedasticity	p = 0.015**	p = 0.831	p = 0.000***	
4. Structural Break	p = 0.288	p = 0.781	p = 0.004***	
'**' = significant at 5 percent ; '***' = significant at 1 percent				

Lastly, we add Bitcoin-related Google Trends and Twitter volume to our model (Table 4). The addition of these two factors does not improve the fit of the model; adjusted R-squared is broadly unchanged and the intercept (alpha) is no longer significant. Removing outliers (results not shown) does not qualitatively change this result.

Table 4. Factor Regressions, with Twitter Factors

	Dependent variable:		
	Equal-Weighted	Market-Weighted	Principal Component
	(1)	(2)	(3)
Mkt_RF	1.617^{***}	1.205^{***}	2.982^{***}
	(0.402)	(0.312)	(0.944)
SMB	0.576	0.724^{**}	0.665
	(0.462)	(0.360)	(1.087)
HML	-0.230	-0.240	0.359
	(0.294)	(0.229)	(0.691)
tweets	-0.00000	-0.000	-0.00000
	(0.00000)	(0.00000)	(0.00000)
google_trends	0.0001	-0.00000	0.0001
	(0.0001)	(0.0001)	(0.0002)
Constant	0.005	0.005	0.005
	(0.008)	(0.006)	(0.019)
Observations	255	255	255
R^2	0.101	0.105	0.053
Adjusted R^2	0.083	0.087	0.034
Residual Std. Error (df = 249)	0.058	0.045	0.137
F Statistic (df = 5; 249)	5.618^{***}	5.853^{***}	2.814^{**}
Note:			*p<0.1; **p<0.05; ***p<0.01

Conclusions

We find evidence consistent with crypto portfolios as “high beta” with respect to market risk, with statistically significant, positive excess return. This evidence suggests that a diversified crypto portfolio can add value to a traditional portfolio of stocks, with a pickup in return.

However, caution is warranted in interpreting the results. The explanatory power of the factor model is low, and no factors other than the market consistently show a statistically significant relationship with our crypto portfolios. We also find mixed evidence on the statistical robustness of the factor regressions. We are also cautious given the large outliers, missing values, and aforementioned presence of survivorship bias in the data—adjusting for survivorship bias by analyzing failed coins would most likely weaken our results.

In a more in-depth study, we would adjust for survivorship bias by running estimations that would include failed coins (perhaps developing an investment rule for when a coin might be added to a portfolio, based on its trading frequency, market depth and liquidity). We might also simulate out of sample performance, forming portfolios in a subsample of the data and running the analysis on a hold-out sample. We might also try adjusting for heteroskedasticity, using robust standard errors. Adjusting for missing values raises further complications, as we would need to avoid imputing prices for days on which the coin in question does not trade for technical reasons (e.g. breakdown in the underlying technology).

Annex

Annex Table 1. Factor Regressions, Outliers Removed

	Dependent variable:		
	Equal-Weighted	Market-Weighted	Principal Component
	(1)	(2)	(3)
Mkt_RF	1.481***	1.182***	2.565***
	(0.414)	(0.322)	(0.969)
SMB	0.714	0.739**	1.002
	(0.467)	(0.363)	(1.091)
HML	-0.173	-0.240	0.434
	(0.292)	(0.227)	(0.683)
VIXCLS	-0.001	-0.0002	-0.003*
	(0.001)	(0.001)	(0.002)
Constant	0.031*	0.008	0.086**
	(0.018)	(0.014)	(0.042)
Observations	255	255	255
R^2	0.099	0.105	0.062
Adjusted R^2	0.085	0.091	0.047
Residual Std. Error (df = 250)	0.058	0.045	0.136
F Statistic (df = 4; 250)	6.874***	7.366***	4.138***
Note:	*p<0.1; **p<0.05; ***p<0.01		

Annex Table 2: Regression Diagnostics-Outliers Removed

1. Serial Correlation	p = 0.070*	p = 0.168	p = 0.264

2. Heteroskedasticity	p = 0.088**	p = 0.815	p = 0.003***
3. Structural Break	p = 0.0571	p = 0.9383	p = .1944
'*' = significant at 10 percent level; '**' = significant at 5 percent level; '***' = significant at 1 percent level			

Bibliography

1. Kash, Top 100 Cryptocurrencies Historical Dataset. https://www.kaggle.com/datasets/kaushiksuresh147/top-10-cryptocurrencies-historical-dataset

2. Pmohun, Complete Historical Cryptocurrency Financial Data. https://www.kaggle.com/datasets/philmohun/cryptocurrency-financial-data

3. https://fred.stlouisfed.org/series/VIXCLS

4. Kenneth R. French, Current Research Returns https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

5. Hileman, Garrick and Rauchs, Michel, 2017 Global Cryptocurrency Benchmarking Study (April 6, 2017). Available at SSRN: https://ssrn.com/abstract=2965436 or http://dx.doi.org/10.2139/ssrn.2965436

6. Source: Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction. Princeton University Press.

7. Brière, M., Oosterlinck, K., & Szafarz, A. (2018). The rise of cryptocurrency Ponzi schemes. Journal of Financial Crime, 25(1), 143-153.

8. Yew, K. T., & Zhang, Y. (2019). Cluster analysis of cryptocurrency price data: Evidence from Bitcoin, Ethereum, Ripple, Litecoin, and Dash. Journal of Risk and Financial Management, 12(4), 180.

9. Abraham, Jethin; Higdon, Daniel; Nelson, John; and Ibarra, Juan (2018) "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis," SMU Data Science Review: Vol. 1: No. 3, Article 1. Available at: https://scholar.smu.edu/datasciencereview/vol1/iss3/1

10. Allen, Hilary (2022), "The Superficial Allure of Crypto," Finance and Development, September 2022. Available at https://www.imf.org/en/Publications/fandd/issues/2022/09/Point-of-View-the-superficial-allure-of-crypto-Hilary-Allen

11. Auer, Raphael; Farag, Marc; Lewrick, Ulf; Orazem, Lovenc; and Zoss, Markus (2022), Banking in the shadow of Bitcoin? The institutional adoption of cryptocurrencies. BIS Working Papers No 1013, 18 May 2022. Available at https://www.bis.org/publ/work1013.htm

12. Benigno, G; and Rosa, C (2023): !The Bitcoin-Macro Disconnect", Federal Reserve Bank of New York Staff Report. Available https://www.newyorkfed.org/research/staff_reports/sr1052

13. Hougan, Matt (2022), "Crypto's Gone Mainstream: Here's What Comes Next." Forbes, February 2022. Available at https://www.forbes.com/sites/matthougan/2022/02/16/cryptos-gone-mainstream-heres-what-comes-next/?sh=140bdfc86432

14. French, Kenneth, and Eugene Fama, Common Risk Factors in the Returns on Stock and Bonds, Journal of Financial Economics Volume 33 Issue 1, February 1993, Pages 3-56.

15. https://bitinfocharts.com/comparison/google_trends-btc.html#alltime

16. https://bitinfocharts.com/comparison/tweets-btc.html#alltime