



HARVARD UNIVERSITY
The Graduate School of Arts and Sciences

DATA SCIENCE II
AC109B

Predicting Cryptocurrency Returns

Students:

Ali Dastjerdi
Angelina Massa
Nate Stein
Sachin Mathur

Professors / Advisor:

Pavlos Protopapas
Mark Glickman
David Wihl

April 27, 2018

Table of contents

1	Introduction	3
1.1	Problem Statement and Motivation	3
1.2	Data	3
2	Literature Review	3
2.1	Momentum	4
2.2	Trading Volume	4
2.3	Covariance between Cryptocurrencies	5
3	Modeling Approach	5
3.1	Baseline Model	5
3.2	d	5
4	Conclusion	5

1 Introduction

Cryptocurrencies are arguably the most polarizing topic within the financial services and financial technology (“fin-tech”) community over the past 5 years. As of April 5, 2018, the market capitalization (“market cap”) of cryptocurrencies was over \$255 billion, a number that easily fluctuates by tens of billions of dollars given the immense volatility associated with cryptocurrencies. Billions of dollars more are being invested in cryptocurrency-related start-ups. As one would expect, those market participants looking to profit from the change in prices of cryptocurrencies have become increasingly active.

1.1 Problem Statement and Motivation

Our goal is to build a predictive model for the price return of cryptocurrencies with the ultimate aim of such a model leading to a profitable trading strategy. As cryptocurrencies become increasingly accepted as financial assets by mainstream investors, the results from these sorts of predictive modeling exercises have significant implications for investors and market-makers. To build this predictive model, features we will experiment with include trading volume and price history of cryptocurrencies, the price histories of other financial assets, and market-signal features we will engineer using natural language processing (NLP) techniques on financial markets news.

1.2 Data

We focus on the following five cryptocurrencies, which, as of 4/1/2018, were among the top 10 cryptocurrencies in terms of market cap on coinmarketcap.com and have data going back to 2015 or earlier: Bitcoin (btc), Ethereum (eth), Ripple (xrp), Litecoin (ltc), and Stellar (xlm). This allows us to have over 950 data points if we decide to use daily rolling returns in our design matrix and use a time frame from 8/8/2015 (the earliest date all five have data available together) to 3/31/2018. Bitcoin and Litecoin have relatively earlier start dates compared to others. Connecting this back to our problem statement, our goal is to build a model that would enable us to predict the price return of one of these five cryptocurrencies by using information on the other four cryptocurrencies.

In addition to using information on the other four cryptocurrencies, we will also use the price returns on other financial assets (such as equity indices, bond prices, foreign exchange prices, gold, volatility indices of exchanges, prices of major commodities, etc.). It is our hypothesis that the correlation with more traditional financial assets will become stronger (although regarding the direction, were not sure about) over the time period analyzed as cryptocurrencies become more mainstream in the investment community.

2 Literature Review

Unfortunately, there is a dearth of sophisticated analysis on cryptocurrency price return forecasting. Although articles with titles such as “Use Machine Learning to Predict Bitcoin” abound, they

are geared more towards an audience interested in learning software engineering / data science by applying a smattering of techniques to an oversimplified version of the problem. To remedy these inadequacies, we aim to contextualize the problem of cryptocurrency price return forecasting within the broader literature on financial time series analysis.

Investment approaches are generally bifurcated between (i) “fundamental” investors, those who value companies based on some estimation of the company’s “true” value based on its assets, profitability, cash flows, etc., and the (ii) “quantitative” (“quant” for short) investors, who use statistical patterns and machine learning techniques to predict future price changes without regard for the “true” value of their investments. With cryptocurrencies, no attempt has really been made to estimate their “true” value, which is not surprising given how relatively new the underlying technology is and to the fact that cryptocurrencies do not lend themselves easily to any sort of parity valuations used with traditional currencies (such as the EUR/USD exchange rate). We focused on “quant” research for stocks and other traditional financial assets with the aim of drawing parallels to the cryptocurrency focus of this paper. The ultimate hope for these endeavors was that they would inspire ideas for feature engineering and other approaches to our problem. Each of the following subsections centers around summarizing a particular quantitative approach and noting how the existing research influenced how we explored our problem.

2.1 Momentum

Momentum is the “tendency of assets with good (bad) recent performance to continue overperforming (underperforming) in the near future” (Vayanos et al 2013). Han, Yufeng, et al (2013) note that this has been one of the most robust empirical tendencies in financial markets, even though it is, in a way, counterintuitive to those subscribing to the “buy low, sell high” mentality. The most common “technical analysis” strategy to profit off this effect is by using a “moving-average” (MA) strategy whereby an investor buys or continues to hold an investment in an asset when the prior day’s price is above its 10-day MA price or invests it into a risk-free asset (such as the 30-day Treasury bill) otherwise. To allow for this effect to potentially be captured by our model, we chose to engineer features capturing whether a given cryptocurrency is above its x -day moving price, where we experimented with several different x -values, i.e., moving-average windows.

2.2 Trading Volume

Vayanos et al (2013) attribute the cause of the momentum effect to some “shock” (e.g., a news event) that impacts the fundamental value of some assets. For example, in February 2018 a Forbes story broke out that major banks, including J.P. Morgan Chase, Bank of America, and Citigroup would no longer allow their customers to purchase cryptocurrencies using their credit cards. This news could be considered a negative shock to the supposed utility of cryptocurrencies as transactional devices and stores of value. Vayanos et al (2013) argue that the momentum effect occurs as a result of the *gradual* outflows that proceed such a news event, preceded by the large sell-off by major investors. Brown et al (2009) further suggest that trading volume could indeed be a proxy

for a number of other important factors, such as liquidity, momentum, and other information. They find that “past trading volume predicts both the magnitude and the persistence of price momentum.” Therefore, trading volume data for cryptocurrencies may enable us to capture momentum and other effects.

2.3 Covariance between Cryptocurrencies

Any trading strategy relying on historical price patterns of multiple assets should be especially concerned with conducting a nuanced analysis of the covariance between those assets. This is potentially the most lacking area in existing literature on cryptocurrency price returns. Yan et al (2017) explore the covariance between cryptocurrencies from a variety of different angles and conclude that the covariance between cryptocurrencies is particularly unstable. This relative instability is true whether we compare cryptocurrencies to a basket of stocks from the same sector (such as a basket of energy stocks) or to a basket of large-capitalization (large-cap) stocks. This is important for our problem in many ways. First, any incorporation of covariance (even if implicit by virtue of including multiple cryptocurrencies in the design matrix) should permit rolling windows to adjust in some form to “learn” any changing comovement patterns. Second, this instability will make it difficult to capture any serial dependence in cryptocurrency returns. DeMiguel et al (2014) use vector autoregressive models to select between equity portfolios and find that the expected returns are greater when the principal components can be used to determine which stocks will experience a positive (negative) return in the future. Work done by Yan et al (2017) suggests this would not be possible with cryptocurrencies, but at the very least, suggests caution when basing a model on any assumption of stable comovement.

3 Modeling Approach

We tried modeling both a quantitative output (regression problem) and

3.1 Baseline Model

3.2 d

4 Conclusion

References

- Brown, Jeffrey H., et al. "Trading Volume and Stock Investments." *Financial Analysts Journal*, vol. 65, no. 2, 2009, pp. 6784.
- Cheng, Evelyn. "JPMorgan Chase, Bank of America & Citi bar people from buying bitcoin with a credit card." *cnn.com*. 2 Feb 2018.
- DeMiguel, Victor, et al. "Stock Return Serial Dependence and Out-of-Sample Portfolio Performance." *The Review of Financial Studies*, vol. 27, no. 4, 2014, pp. 10311073.
- Han, Yufeng, et al. "A New Anomaly: The Cross-Sectional Profitability of Technical Analysis." *The Journal of Financial and Quantitative Analysis*, vol. 48, no. 5, 2013, pp. 14331461.
- Longmore, Kris. "A Framework for Applying Machine Learning to Systematic Trading." *Quantify Partners*. RobotWealth.com.
- Vayanos, Dimitri, and Paul Woolley. "An Institutional Theory of Momentum and Reversal." *The Review of Financial Studies*, vol. 26, no. 5, 2013, pp. 10871145.
- White, Halbert. "A Reality Check for Data Snooping." *Econometrica*, vol. 68, no. 5, 2000, pp. 10971126.
- Yan, Yihang, et al. "Do cryptocurrencies move in a parallel manner?." *Harvard University Project*, 2017.