Principal Component Analysis of Cryptocurrencies: Implications for Investors ........ 2
Some Questions ........ 2
Introduction ........ 2
Methodology ........ 3
Motivation for PCA ........ 3
Appropriateness of PCA ........ 3
Past PCA Literature ........ 3
Implications ........ 4
Can Cryptocurrencies be Viewed as an Asset Class? ........ 4
Do PCA Results Differ Significantly over Different Time Periods? ........ 4
References ........ 6

# *Principal Component Analysis of Cryptocurrencies: Implications for Investors*

## Some Questions

- Finalize the frequency over which we're going to compute rolling returns (e.g., daily, weekly, monthly). Most financial research uses monthly or longer, but depending on how the numbers shake maybe that influences what we do? Alternatively, maybe it's easy to show the results for several frequencies and then pick one as the point of reference for rest of analysis.
- In order to determine whether the price returns of cryptocurrencies move in a parallel manner, what should we compare the crypto PCA results with? I give the results of two studies below that asked the same question (one on Dow Jones stocks and the other on treasuries), but it would probably be helpful to compare to one or two other subsets of assets that we analyze ourselves.
    - Just a thought: maybe one of those asset groups could be the top holdings in the XLE energy ETF? https://us.spdrs.com/etf/energy-select-sector-spdr-fund-XLE. I think of all the ETFs to choose from that we want to represent as being a strong sampling of assets representing an industry, it would be energy. Then we could compare PCA on these energy stocks vs. the cryptos.
    - Nice place to download stock data besides Yahoo:
        - http://www.nasdaq.com/quotes/historical-quotes.aspx.
- Should we have one or two paragraphs summarizing Bitcoin? Kind of like the material in the PDF Yihang sent *The Cross-Section of Crypto-Currencies as Financial Assets - An Overview*.

## Introduction

This paper examines the statistical distribution of the price patterns of cryptocurrencies through the lens of principal component analysis (PCA). Although PCA is routinely applied to studies of traditional asset classes and incorporated in various econometric models, it has yet to be applied in a significant manner to cryptocurrencies. The rise of cryptocurrencies is the most intensely debated paradigm shift in financial markets this past decade, with opinions ranging from JPMorgan CEO Jamie Dimon calling them a "fraud" to Bill Gates claiming "Bitcoin is better than currency." Moreoever, cryptocurrencies have only recently entered a new stage in their history: trading on recognized exchanges. The Chicago Board Options Exchange (CBOE) debuted bitcoin futures in the U.S. market on December 10, 2017, with the immense surge in activity triggering multiple trading halts designed to stablize the marketplace. Aside from a brief introduction to cryptocurrencies, a discussion on the merits and future of these innovations is beyond the scope of this paper.

Instead, we will assume cryptocurrencies are not going away anytime soon and focus our analysis on other issues pertinent to potential investors and market-makers in cryptocurrencies. **First, we will use PCA to analyze whether the price returns of cryptocurrencies move in a parallel manner**. By comparing the comovement of cryptocurrencies with the comovement of other asset classes, one can more methodically evaluate the degree to which cryptocurrencies truly represent an asset class. **Second, we sample different time frames and repeat the analysis to determine whether there have been any meaningful structural shifts in the cryptocurrency market.** Finally, we highlight several implications our results have for portfolio construction, risk management, and trading. In that section, we will offer an opinion over whether cryptocurrencies could be viewed as an asset class.

## Methodology

### Motivation for PCA

Before delving into the specifics of this paper's implementation of PCA, let us motivate the benefits for studying asset returns with PCA. PCA provides a method with which to reduce the complexity of a large data set of asset price returns. For the purposes of this study, it enables the identification of the underlying statistical factors that cause the comovement in cryptocurrency returns. Each of these factors aims to account for the greatest degree of variation in the data as possible and is orthogonal (i.e., uncorrelated) to the other factors. These findings can be used to quantify the relative importance of each cryptocurrency in explaining the movement of the cryptocurrency market as a whole.

The first principal component is the component that *best* explains variation in the underlying data, i.e., the greatest amount of variation, and is of particular importance to this study. In finance, risk is frequently broken down into two categories: *systematic* (i.e., *market*) risk that can be mitigated through a diversified portfolio and *idiosyncratic* risk that is specific to each individual asset and cannot be diversified away. In applications of PCA on asset returns, the first principal component is generally accepted as a representation of the overall return of the assets, arguably representing the return for taking on the systematic risk of those assets. In other words, if all the assets shared the same idiosyncratic risks, the first principal component could be conceptualized as the "equal weighted market index" (Shukla and Trzcinka, 1990). Thus, for assets that are fairly correlated with one another, i.e., a large proportion of their comovement is accounted for by the general fortunes (or misfortunes) in their market, one would expect the first component to account for a relatively large proportion of variance and to have similar loadings on the variables.

### Appropriateness of PCA

One of the basic assumptions behind using PCA is that the variables being examined have a strong linear relationship with each other; else, it would be unlikely that they share any meaningful common components. To assess the strength of the variables' linear relationship and, consequently, suitability for PCA, one can use the correlation coefficients between the pairs of the variables (Hair, 2010). Here we illustrate the correlation matrix of the cryptocurrency [monthly] returns.

[Placeholder]

| | rip | btc | eth | ltc | mon | nem | dash |
|---|---|---|---|---|---|---|---|
| **rip** | 1.000 | 0.305 | 0.325 | 0.694 | 0.066 | 0.630 | 0.037 |
| **btc** | 0.305 | 1.000 | 0.164 | 0.322 | 0.102 | 0.349 | 0.031 |
| **eth** | 0.325 | 0.164 | 1.000 | 0.192 | 0.171 | 0.361 | 0.470 |
| **ltc** | 0.694 | 0.322 | 0.192 | 1.000 | 0.044 | 0.428 | 0.018 |
| **mon** | 0.066 | 0.102 | 0.171 | 0.044 | 1.000 | 0.114 | 0.128 |
| **nem** | 0.630 | 0.349 | 0.361 | 0.428 | 0.114 | 1.000 | 0.133 |
| **dash** | 0.037 | 0.031 | 0.470 | 0.018 | 0.128 | 0.133 | 1.000 |

[Short discussion]

### Past PCA Literature

Broadie's (2012) PCA on very similar assets, on-the-run Treasury securities of varying maturities, demonstrates an example of how the first component accounts for a large degree of variation (95%) and the loadings of the first component are near-equal for each Treasury security. In short, treasuries do demonstrate a strongly parallel price shift. Of course, it is difficult to posit theoretically or empirically that treasuries of varying maturities would have significant idiosyncratic risk. Their greatest (shared) risk, or opportunity depending on whom you ask, is the risk of changing interest rates. Feeney and Hester (1967), in one of the first studies of its kind, conduct PCA on the rates of return of the 30 Dow Jones industrial (DJI) stocks over a 50 quarter period and find that the **first component accounts for 41% of the variance** (the second component accounts for 9%) with all loadings being positive for the first component.

## Implications

### Can Cryptocurrencies be Viewed as an Asset Class?

In equity markets, the question arises as to whether it is useful to conceptualize certain stocks as belonging to an industry. For example, let's presume one of the so-called FANG[1] stocks, Facebook, reports strong quarterly profits and consequently its stock price appreciates significantly, yet it reports no specific news accounting for its performance. Is its quarterly success the result of it enjoying a competitive advantage among other similar companies, or is it that all of the stocks in the technology sector are generally doing well? A test of whether it is fair to look at a set of stocks belonging to an industry is whether those stocks tend to load similarly on the different components of that industry.

We apply this approach to the cryptocurrency market. If the price returns of cryptocurrencies are highly correlated, then investing in more than one cryptocurrency provides no advantage from the perspective of building a diversified portfolio to reduce overall variance (unless, perhaps, you were to "short" one or more of the other cryptocurrencies).

**[To do]** Perform t-test done in Feeney and Hester (1967)

> Null hypothesis: Correlations of two cryptocurrencies with a principal component will be of the same sign with probability 0.5. By examining the frequency with which a pair of stocks have the same sign of correlation coefficient with a number of different components we may test the hypothesis underlying the aggregation of stocks into industry classifications.

### Do PCA Results Differ Significantly over Different Time Periods?

**For an investor to use the findings from the previous section in making portfolio allocation decisions, an implicit assumption is that the covariance matrix will continue relatively unchanged into the future**. **This question is relevent whether or not the prior section determines that cryptocurrencies can be fairly viewed as an asset class because that opinion should be contextualized within the continuity of the covariance matrix.**

Whether PCA results are similar or not over different time frames, an investor will want to keep this assumption salient in their planning decisions. However, if this continuing covariance matrix assumption does not even appear to be supported in the time frames we analyzed, then serious caution should be exercised in basing any investment decisions off it.

To analyze these, we examine three properties of the PCA results over subperiods in our original timeframe:

- Are loadings for each of the cryptos the same when estimated from each subperiod?
  - How many of the loadings (weights) change signs between two subperiods?
- Is the percentage of explained variance estimated from each subperiod the same?
- Can components estimated in an earlier subperiod help us predict crypto price movements in the next period?

---

[1] Facebook, Amazon, Netflix and Google.

**From** ***Asset Returns and the Listing Choice of Firms (Baruch and Saar) [2009]*** (p. 2261)

Their test (2260):

> *We want to look at the "distance" of the switching portfolio's loading from the loadings of the NYSE and NNM portfolios, and how this distance changes over time. We therefore compute 30 distances of the switching portfolio's loading from NYSE stocks by taking the absolute value of the difference between the loading (say, on the first principal component) of the switching portfolio and the loadings on the same principal component of the 30 NYSE portfolios.*

**Is this an opportunity to incorporate vector norms by taking the difference in loading vectors into our analysis?!?!?**

**Table 4**
**Switching sample: principal component analysis**

Panel A: analysis of loadings on 1st principal component

| Distances | Year $t-2$ \|NYSE-switching\| | Year $t-2$ \|NNM-switching\| | Year $t-1$ \|NYSE-switching\| | Year $t-1$ \|NNM-switching\| |
|---|---|---|---|---|
| Mean | 0.2274 | 0.1735 | 0.1894 | 0.2459 |
| Median | 0.1975 | 0.1224 | 0.1883 | 0.2688 |
| Standard deviation | 0.0903 | 0.1174 | 0.0920 | 0.0664 |
| $p$-value of $t$-test (9000 Obs.) | < 0.0001 | | < 0.0001 | |
| $p$-value of $W$-test (9000 Obs.) | < 0.0001 | | < 0.0001 | |
| Avg. of 300 $t$-stat. (30 Obs.) | 7.57 | | −6.10 | |
| $p$-value of average $t$-stat. | < 0.0001 | | < 0.0001 | |
| Avg. of 300 $W$-stat. (30 Obs.) | 2.24 | | −2.14 | |
| $p$-value of average $W$-stat. | 0.0249 | | 0.0323 | |

## ***References***


Baruch, Shmuel, and Gideon Saar. “Asset Returns and the Listing Choice of Firms.” *The Review of Financial Studies*, vol. 22, no. 6, 2009, pp. 2239–2274. *JSTOR*, JSTOR, www.jstor.org/stable/30225715.

Shukla, Ravi, and Charles Trzcinka. “Sequential Tests of the Arbitrage Pricing Theory: A Comparison of Principal Components and Maximum Likelihood Factors.” *The Journal of Finance*, vol. 45, no. 5, 1990, pp. 1541–1564. *JSTOR*, JSTOR, www.jstor.org/stable/2328749.

Hair, Joseph F., et al. *Multivariate Data Analysis*. 7th ed., Pearson Education Limited, 2010.

Hester, Donald D, James Tobin, and George J. Feeney. *Risk Aversion and Portfolio Choice*. New York: J. Wiley, 1967.