TALOS.

ANALYSIS

Stylized Facts for Cryptocurrencies: A Sectoral Analysis

STYLIZED FACTS FOR CRYPTOCURRENCIES:

A SECTORAL ANALYSIS

Sonakshi Rohra and Ilya Kulyatin, Cloudwall* Research

ABSTRACT: Our last paper briefly studied the statistical properties of Bitcoin's price process. In this paper, we take a step forward and make a thorough study of stylized facts across the crypto market. The paper begins with a statistical analysis of a top 10 token - market cap weighted index. Next, we use the DACS taxonomy to construct sectoral indices and compare statistical behaviour across the 5 sectoral indices. A range of techniques and metrics are utilized such as distribution fitting, Q-Q plots for distribution comparisons, skewness, kurtosis, GARCH fitting and autocorrelation plots to confirm or deny the common stylized facts.

INTRODUCTION

Although it is hard for us to generalize the behavior of time series, there are usually some **statistical properties** inherent to financial data. These properties are known as **stylized facts** (Cont, 2001) and are the result of years of empirical studies about the statistical behavior of financial asset prices. By definition, these stylized facts are common across financial markets, and we want to uncover them also in cryptocurrencies. To keep our analysis empirical, here we use non-parametric methods to evaluate our time series, i.e. we do not make any assumptions, if not when clearly stated otherwise.

DATA

We will be using a limited time-series dataset selected from the top **50 tokens** by market cap, with daily frequency and sampled between **September 2020 and March 2022**. The source is CoinGecko, freely available and thus easy to replicate. The market cap information is from CoinMarketCap. This market cap is the product between price and the circulating supply. While there are better ways of representing the size of a project, selecting a more appropriate metric is out of scope here. The data sample is short, as

^{}Cloudwall and the technology behind its Serenity System were acquired by Talos in April 2024.*

most of the tokens relate to protocols launched in the past couple of years, but it's sufficient for an initial analysis.

First, we extract the top 10 coins by market cap, naming them the Top 10 Index: Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), Ripple (XRP), Terra (LUNA), Cardano (ADA), Solana (SOL), Avalanche (AVAX), Polkadot (DOT) and Polygon (MATIC). For reference, we list the starting date of the available data for each token in the top 10:

List of the selected Top 10 tokens with the respective start dates.

The final date range is **22 September 2020 to 25 March 2022**, as this allows us to include AVAX. We plot the price data for the 10 tokens below, rebasing the dataset and log transforming for easier visual analysis.

g.

When carrying out statistical analysis of a time series, we do want to see **stationary data**, which means its stochastic properties are invariable with respect to time. We want this because otherwise the statistical methodologies usually applied will show misleading ("spurious") results. Most notably, nonstationary variables have **infinite variance**, which is problematic in our Gaussian financial world. For those interested to know where these stationarity considerations started, check out Box and Jenkins (1970), or something more modern, like the outstanding introductory Econometrics manual by Kennedy (2008). To get closer to stationarity, let's take the first difference in prices.

LOG RETURNS VS SIMPLE RETURNS

There are two main ways we get these first differences: using log-returns (L_t) or simple returns (S_t) :

$$
L_t = \ln\left(\frac{P_t}{P_{t-1}}\right)
$$
 $S_t = \frac{P_t - P_{t-1}}{P_{t-1}}$

For our analysis we prefer using log returns for a couple of reasons:

- Firstly, log returns fit a normal distribution better than simple returns, and thus they fit the assumptions of most stochastic pricing models much better. Not surprisingly, financial prices are usually not Normal, but log-Normal (it's indeed pretty rare to see negative prices), so we expect the log return to be close to Normal.
- The theoretical range for log returns is *(-inf, +inf)*, and thus symmetric around 0, whereas simple returns have a lower bound of -100% (when *P_t = 0*, and

ø

assuming prices are not negative). And we do like continuous returns, it makes a quant's life easier.

And this is how it looks:

Here we will also be analyzing a market-cap-weighted index for these Top 10 tokens, so here's a table with the used weights (as of March 2022, taking the average market cap for the past 30 days):

For a more realistic weighting, it's better to take into account the **free float**, excluding e.g. lost wallets or locked tokens. For the sake of simplicity, we use the fully diluted market cap, but in the future, we'll share results with more realistic supply metrics.

Ø

It should be noted that for simple returns, the index return is obtained by using a dot product of the weight vector with the simple returns vector. However, in the case of the log returns, these returns are *cumulative*, and so we cannot obtain the index return using a dot product. In this case, we use simple returns to compute the index return and take the log of these cumulative simple returns.

$$
r_{p,L} \neq \sum_{i=1}^{n} w_i r_{i,L}
$$
 where $r_{i,L} = \log$ return of asset i

STYLISED STATISTICAL PROPERTIES OF THE TOP 10 INDEX

Here we will refer to the same **stylized facts** as outlined in Cont (2001). You can also check our earlier review of a paper on BTC stylized facts.

1. **Absence of autocorrelations:** Autocorrelations of asset returns are often insignificant, except for intraday time scales (~20 mins) for which microstructure effects usually come into play. As higher frequency data is out of scope for this study, looking at daily returns we expect to see no autocorrelations, i.e., our data is *stationary*. Here's an autocorrelation function (ACF) for the returns of our Top 10 Index:

As expected, we notice how the autocorrelation is significant for lag 0, but negligible for higher lags, indicating that our percentage change in index prices is independent between the adjacent periods. Ljung-Box procedure confirms these results.

This means that our price series cannot be accurately predicted by just using a univariate time series forecast. Bummer, no easy money, the moon will have to wait.

2. **Aggregational Gaussianity:** It is usually observed that, as we increase the time scale $delta_t$ over which returns are calculated, they get closer to a Gaussian distribution. First, we check whether the daily returns $(delta_t = 1 day)$ show any fit for the Normal distribution. We use a histogram to aggregate our returns data into a set of bins and fit a Normal distribution with the historical mean and standard deviation with empirical parameters $mu = 0.003601$ and $sigma = 0.0378368$.

From a preliminary observation, we notice that while the data approximately resembles *some* bell curve, a Normal distribution with the above parameters Ø

 $(mu, sigma)$ may not provide the best fit since the histogram does not seem symmetric and is heavier tailed than a Normal. We can further analyze Normality by observing the Q-Q plot of the data. Note that the Q-Q plot maps the quantiles of the empirical returns to the quantiles of a theoretical Normal distribution. If our empirical data is Normally distributed, we should expect to see a linear mapping to the theoretical Normal quantiles:

The Top 10 Index returns look positively skewed. In stock markets, we usually observe positive skewness for single names, while for aggregates we see a negative skewness. Here maybe 10 assets are not enough to move the skewness (and the index is also 50% in BTC).

We usually refer to the Prospect Theory to explain why investors like positively skewed returns. For a brief and insightful discussion on the skewness in financial returns, check out this piece by Man Institute.

Visual analysis is not enough, so let's use the Shapiro-Wilk test for Normality. If you wonder why Shapiro-Wilk vs other tests, then check the small sample size issues with these types of tests. This is the test result for our Top 10 index:

Note that since the p-value is extremely low, we can conclude that indeed, our data is not Normally distributed.

We note that while it would be interesting to observe whether coarse-grained data (higher *delta_t*) would produce a Gaussian fit, this dataset is not enough to observe a distribution at much higher order *delta_t*, say monthly data, as that would give us only about 18 data points to work with. However, we can study the returns for every 3 days, and as suspected, this aggregational behavior shows statistically significant results for Normality.

We take the price data every 3 days and compute the percentage change. Note that we have about 183 data points for the 3 days returns, so definitely sub-optimal, but better than nothing.

We note the results of the Shapiro-Wilk test for the 3-day simple returns:

Note that the p-value for the test statistic is slightly above 0.05, so we can say that here the Normality hypothesis is not rejected at least at the 10% confidence level. Great, we've shown that there indeed seems to be aggregational Gaussianity.

This result is significant as it highlights the importance of selecting delta_t such that our data conforms with the usual Normality assumptions for statistical modelling

purposes. For investors, this means that different risk time-frames require model adjustments to avoid biased results.

3. **Heavy Tails:** The unconditional distribution of returns usually displays a power-law or Pareto-like tail, with a tail index that is finite, higher than two, and less than five for most datasets studied. The easiest way to check for tails of an empirical distribution is using the fourth moment, kurtosis, or equivalently excess kurtosis:

$$
\text{kurtosis }\kappa = \frac{[X-\mathbb{E}(X)]^4}{\sigma^4}, \quad \text{excess kurtosis } K = \frac{[X-\mathbb{E}(X)]^4}{\sigma^4} - 3
$$

We note the results for the kurtosis of the Top 10 Index:

The negative excess kurtosis confirms our hypothesis about a light-tailed distribution from our Q-Q plot. Thus, we do not have heavy tails for the log-returns of our Top 10 Index, which is the first sign of diversification benefits.

In contrast, if we look at the kurtosis for log-returns of an individual token, ETH, for example, we have results indicating a great tendency of a heavy-tailed distribution:

Ø.

Note that the excess kurtosis for the Top 10 Index, while negative, is much closer to 0 than the excess kurtosis of ETH, i.e. it's closer to a Normal, which means that such an index will generally perform at a stable rate, with chances of extreme returns close to what we would see with a Gaussian distribution.

For investors this means that a cap weighted index seems to be diversifying away much of the heavy tails associated with investing into single name tokens. Will this have an impact on the pricing of derivatives on baskets of tokens? We are gearing up our derivatives capabilities and will be sharing interesting results in the coming months. Let us know if that's something you are interested in.

4. **Gain and loss asymmetry:** Cont (2001) states that there is an asymmetry in positive and negative returns, with the latter being larger. This may be confirmed by taking some returns data for any preferred large-cap equity and observing the skewness of the data, which is most likely to be positive under stable market conditions. Additionally, we may point to Albuquerque (2012) for further discussion on differences in skewness for aggregate and individual stock returns. This positive skewness on individual assets implies that an investor may expect frequent smaller losses and a few large gains, thus attaching more risk to their investment.

However, we note that in our case, the skew for individual tokens for our data is negative, indicating that the investor may expect frequent smaller gains and a few larger losses on their investment. As a curiosity, this is one of the potential behavioral explanations for the Momentum factor.

An easier way to analyse this asymmetry without using statistical jargon would be to corroborate our expectation of 'frequent smaller gains and few larger losses' by simply looking at the count of our negative returns and positive returns, and the maximum and the minimum of our returns distribution.

We can easily see that while the frequency of negative returns is lower, the absolute value of our minimum returns is larger. On the other hand, we have a greater frequency for positive returns, although the absolute value of our maximum returns is lower.

Thus, it must be noted that while such an index may provide more stable profits and attract more risk averse investors, there is still some chance of large losses and thus requires appropriate risk management, perhaps starting with a VaR model with particular attention to tail risks.

5. Returns intermittency and volatility clustering: Usually, empirical returns display a high degree of variability. This is quantified by the presence of irregular bursts in the time series of a wide variety of volatility estimators. We may observe this by studying the standard deviation of our index returns on a 20-day rolling basis:

We note that the rolling standard deviation shows some irregular patterns at random, such as the large rise in volatility around June 2021, followed by a dip in August 2021, and a small rise again starting around September 2021.

At first glance, the behaviour is not seasonal by any means, nor does it show any visible trend as such. We will not cover seasonality or trend analysis today.

Another significant stylized fact that we observe from our Top 10 index is that our volatility measure displays a positive autocorrelation over several days (here, approximately 18 days). This statistically shows the (stylized) fact that indeed highvolatility events tend to cluster in time.

6. Slow decay of autocorrelation in absolute returns: Related to the above, we can see a long-ranged dependency in volatility. This means that large changes in returns are likely to be followed by other large changes, while small changes are followed by other small changes. This can be shown by positively autocorrelated and slowly decaying absolute or squared returns. And this is exactly what we see:

œ

7. **Conditional Heavy Tails:** It is well known that instances of stochastic volatility such as volatility clustering may be modelled by GARCH. Here, we fit a GARCH(1,1) model to our data. We are now interested in the residual time series after we fit the GARCH(1,1) model. Cont (2001) states that even after correcting returns for volatility clustering, the residual time series still exhibit heavy tails, though less heavy than with the unconditional distribution. If our residuals are truly random and the GARCH model accounts for all stochastic behaviour of volatility, we expect that these returns will follow a Normal distribution with zero mean and some constant standard deviation *sigma*, i.e.:

$$
\epsilon \sim N(0, \sigma^2)
$$

It is easy to see whether our **GARCH residuals exhibit heavy tails** using a QQ plot. Note that we removed an outlier from the data by just filtering anything above 3 standard deviations. It's a simplification, be careful using this in your production risk systems.

Lots of these outliers in crypto were actually not outliers when they happened, even *though with a maturing market and a better understanding of the fair value of crypto (if any), we should expect not to see such outliers in the future. This is not a*

Ø

financial advice, but it's extremely interesting to think how these stylized facts will change with the maturity of this market.

Just as in Cont (2001), we observe that the empirical quantiles lie below the normal quantiles for the left tail, and above the normal quantiles for the right tail, thus exhibiting heavy tails. This has **considerable implications for the risk management** of cryptocurrency portfolios.

We'll leave some stylized facts for future analysis, especially around the **leverage effect**, **volume vs volatility correlation**, and **asymmetry in time scales**.

Here's a summary table for the moments of our Top 10 Index and its components:

g.

ANALYSIS BY SECTOR

Now that we've done an initial stylized facts analysis on a diversified index of top 10 tokens, we'd like to see the same about **sectoral indices**. We divide the major 50 tokens by sector based on the DACS taxonomy: **Computing**, **Currency**, **DeFi**, **Smart Contract Platform,** and **Entertainment**. Then we construct sectoral indices of the top 5 tokens for each sector, weighted by their market cap. Lastly, we also create an equal-weighted index out of the five sectoral indices and analyze the stylized facts for each of these derived indices. The division of tokens between sectors is given in the table below (note that we exclude stablecoins from our analysis, for obvious reasons, and we probably should have excluded the *meme coins*):

While we chose to implement a Top 5 Index, notice that the Computing sector consists of only 4 tokens and the Entertainment sector consists of only 3 tokens, so this will impact our analysis, but in a future report, we'll make sure to have more equilibrated sectoral composition. What follows is the distribution of weights for each sector.

SECTOR INDICES: LOG RETURNS

The log returns for each sector are plotted below. To a trained eye, the Entertainment sector looks much more volatile, but otherwise, there's nothing else we can tell from just this visual analysis if not that these indices do look quite correlated (cross-sectional analysis is not part of this report, so we'll not talk about the correlations). In this section,

we'll also use the equally-weighted sectoral index, where we assign 20% to each of the above-mentioned sectors (let's call it EWSI — Equally Weighted Sector Index).

STYLISED STATISTICAL PROPERTIES OF SECTORAL INDICES

1. **Absence of autocorrelations:** We observe the autocorrelation for our daily logreturns data for all of the Top 5 indices:

While the ACF for the index returns dies down quickly after lag 1 for all sectors, there are **a few lags at which the ACF is not insignificant** (at the 95% level), and thus there may be some predictability from past returns — especially for the Computing, DeFi and the Smart Contract Platform indices.

A suggestion to investigate this would be to fit an ARMA model to the returns for each sectoral index, and check the fit of such models.

œ

Note that the equal-weighted sector index also shows slightly significant autocorrelation at lags of 4 and 5, which is unsurprising given the consistent autocorrelation at higher lags in each of the sector indices. We'll eventually look deeper into this.

2. Aggregational Gaussianity: We check whether the daily returns $delta_t =$ *1* day) show any fit for a Normal distribution. We use a histogram to aggregate our returns data into 80 bins and fit a Normal distribution with the historical mean and standard deviation with empirical parameters as follows:

And this is how the PDF looks for the equally weighted index:

œ

That lonely guy in the left tail is the 19–20 May 2021 China crackdown on crypto. Do you still feel the pain?

Next, we can test for Normality for our daily returns data for each sectoral index using the Shapiro-Wilk test:

Since our p-values are very small, we reject the null hypothesis and thus conclude that our data is not Normally distributed. Here we also apply the Shapiro-Wilk test to simple returns over 3-day periods (i.e. *delta_t = 3* days). This is too short of a time frame and we thus would need to explore further this particular stylized fact.

Computing is the most "tame" of the indices, and the other sectors seem to be converging to it in terms of the statistical properties we are talking about here. Why is that?

g.

3. Heavy Tails: Here's a table for the kurtosis of each index (I wonder which two tokens are "dogging" the Currency sector…):

As expected, we notice a positive excess kurtosis, indicating that the returns for all sector indices show heavy tails, although the Currency sector shows the heaviest kurtosis — possibly attributed to the large weight of BTC which no doubt has some pretty heavy tails.

g)

This highlights how important it is to not stop at the "boomer" coins (i.e. BTC and ETH).

4. **Gain and Loss Asymmetry:** We can now check the asymmetry of the returns distribution by reporting the skewness. As noticed earlier, the Computing sector seems to behave like the average of the others, so EWSI does look similar to this sector in terms of its statistical properties.

Given that there's still this negative skewness, in case we want to use this as an investable index it might make sense to apply volatility scaling, which can potentially even reverse the negative skew.

5. **Returns Intermittency and Volatility Clustering:** We compute the 20-day rolling standard deviation to observe volatility over time for each index. Notice how an equally weighted index of these category sub-indices offers a less volatile profile. That's part of an initial hint for the sectoral diversification benefits.

We see a positive, slowly decaying volatility autocorrelation for all sectors, just as we'd expect. Here we are showing only the EWSI index, as you are probably already fed up with all of our charts:

SUMMARY STATISTICS: SECTORAL INDICES AND EWSI

œ

CONCLUSIONS

This concludes our exploratory data analysis around the stylized facts for cryptocurrency returns. We have seen that the traditional finance stylized facts apply to a broad range of cryptocurrencies and that there appear to be diversification benefits from sectoral diversification (at least in our quest for more Normal returns). Stay tuned for further sectoral analysis in the coming weeks, especially around the portfolio diversification insights.

AUTHORS

- Sonakshi Rohra, Quantitative Researcher @ Cloudwall Capital.
- Ilya Kulyatin, Head of Research @ Cloudwall Capital.

DISCLAIMER

Not a financial advice, solicitation, or sale of any investment product. The information provided to you is for illustrative purposes and is not binding on Cloudwall Capital. This does not constitute financial advice or form any recommendation, or solicitation to purchase any financial product. The information should not be relied upon as a replacement from your financial advisor. You should seek advice from your independent financial advisor at all times. We do not assume any fiduciary responsibility or liability for any consequences financial or otherwise arising from the reliance on such information. You may view this for information purposes only. Copy, distribution, or reproduction of all or any portion of this article without explicit written consent from Cloudwall is not allowed.

^{}Cloudwall and the technology behind its Serenity System were acquired by Talos in April 2024.*

talos.com

Disclaimer: Talos Global, Inc. and its affiliates ("Talos") offer software-as-a-service products that provide connectivity tools for institutional clients. Talos does not provide clients with any pre-negotiated arrangements with liquidity providers or other parties. Clients are required to independently negotiate arrangements with liquidity providers and other parties bilaterally. Talos is not party to any of these arrangements. Services and venues may not be available in all jurisdictions. For information about which services are available in your jurisdiction, please reach out to your sales representative. Talos is not is not an investment advisor or broker/dealer. This document and information do not constitute an offer to buy or sell, or a promotion or recommendation of, any digital asset, security, derivative, commodity, financial instrument or product or trading strategy. This document and information are not intended to constitute investment advice or a recommendation to make (or refrain from making) any kind of investment decision and may not be relied on as such. This document and information are subject to change without notice. It is provided only for general informational, illustrative, and/or marketing purposes, or in connection with exploratory conversations with institutional investors and is not intended for retail clients. The information provided was obtained from sources believed to be reliable at the time of preparation, however Talos makes no representation as to its accuracy, suitability, non-infringement of third-party rights, or otherwise. Talos disclaims all liability, expenses, or costs arising from or connected with the information provided.

© Copyright 2022 | Talos Global, Inc.