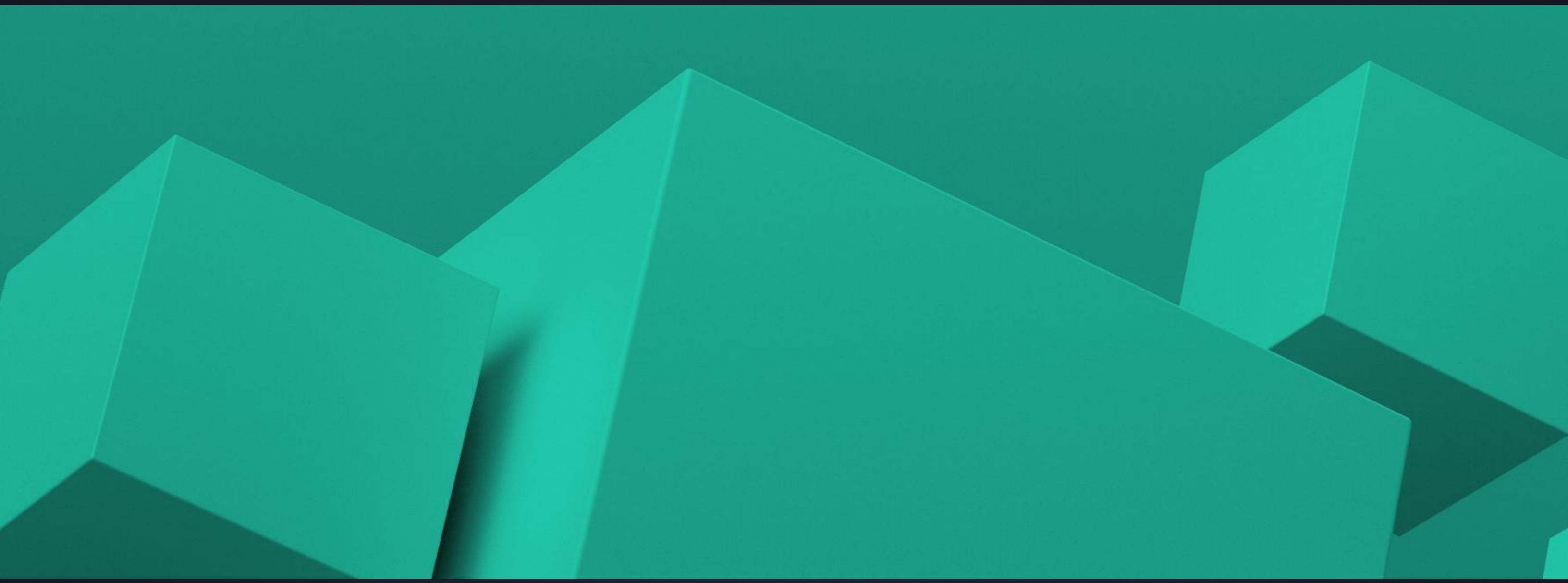
TALOS

ANALYSIS

Risk Models for Crypto Assets: Fundamental vs. PCA



April 2023

Serenity* Risk Models for Crypto Assets:

Fundamental vs PCA

Chunlei Xia
April 18, 2023

Abstract: This report compares factor models and Principal Component Analysis (PCA) for building risk models for digital assets. We explore the benefits of PCA, such as orthogonal factors and reduced over-fitting, as well as its potential to serve as an alternative to factor models. We provide an overview of multifactor models and describe two models that we built at Cloudwall.* We present the PCA model's R-squared by components and compare the first component to market return. Additionally, we compare the in-sample and out-of-sample performance of both models and present bias statistics. Overall, our findings suggest that principal component analysis (PCA) is a promising approach for constructing risk models for digital assets compared to factor models.

MULTI-FACTOR MODELS(MFMS)

Multifactor models identify common factors among assets and provide sensitivities of asset returns to these factors. Such models are helpful tools for modern portfolio risk management. Using a K-factor model (a model that includes K common factors), the excess return of an asset could be represented as:

$$r_{i,t} = \sum_{k=1}^{K} \beta_{i,t}^{k} f_t^k + \epsilon_{i,t}$$

where

$\beta_{i,t}^{k}$ is the risk exposure of asset i to factor k

f_t^k is the return of factor k

$\epsilon_{i,t}$ is the specific return of asset i.

Portfolio risk, as measured by variance, can be easily obtained under the MFM framework (please refer to [1] for more details). This is because we do not need to evaluate the covariance for each pair of assets.

$$\text{risk} = X \Sigma X^T + \Delta$$

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

where

X is the exposure matrix

Σ is the factor covariance matrix

Δ is the specific risk

TYPES OF MULTI-FACTOR MODELS

When evaluating asset market risk, three types of multifactor models are frequently utilized: fundamental, macroeconomic, and statistical factor models. These models vary in their factor construction. Fundamental factor models use observable asset attributes (e.g. market capitalization, price-to-earnings ratio for equity markets), while macroeconomic factor models use observable economic time series (e.g. interest rates, inflation). Statistical models, on the other hand, derive factors from the covariance matrix of asset returns using statistical methods. Fundamental models provide better explanation of factors because they decompose risk using fixed factors that are intuitive and well-understood. There is extensive literature on common fundamental risk factors in cryptocurrency (see reference [2-4]), our current factor model adopts five fundamental factors. Statistical factors are less intuitive, but they are not limited by a fixed factor structure, allowing for better adaptation to evolving market conditions with more responsive factors. Therefore, we expect statistical models to have higher R-squared and more significant factors compared to fundamental factor models.

We are leaving out the macroeconomic factor model at this point. Some research reports suggest an increased correlation between the crypto market and equity market, while others, such as reference [5], have shown that Bitcoin is orthogonal to monetary and macroeconomic news.

We now examine the fundamental model and statistical model more closely.

FUNDAMENTAL MODELS

The first risk model that Cloudwall built is a fundamental model: the Serenity Factor Risk Model (SFRM). This is a crypto factor model which allows to decompose and attribute risk exposures taken by any crypto portfolio or crypto investment strategy. It decomposes asset returns across a number of “style” crypto factors, such as Momentum, Volatility, Size, Market, Liquidity. There are two approaches when building fundamental models:

1. start with constructing factor exposures from observable asset specific fundamentals (commonly referred as Barra approach),
2. use a two-pass regression that starts with constructing factor returns (aka the Fama-French approach).

The Serenity Factor Risk Model uses the second approach.

PCA MODEL R^2 BY COMPONENTS

Given the appealing features of PCA mentioned earlier, we have built a statistical model based on this approach. When studying PCA models, a first metric to consider is the R-squared value of each principal component. Similar to equity market models, the first principal component has by far the largest explanatory power (with an R-squared value of 0.494), and the first five components together have a combined R-squared value of 0.683. For the first version of our PCA model, we have chosen to use five components to facilitate comparison with our current five-factor SFRM model.

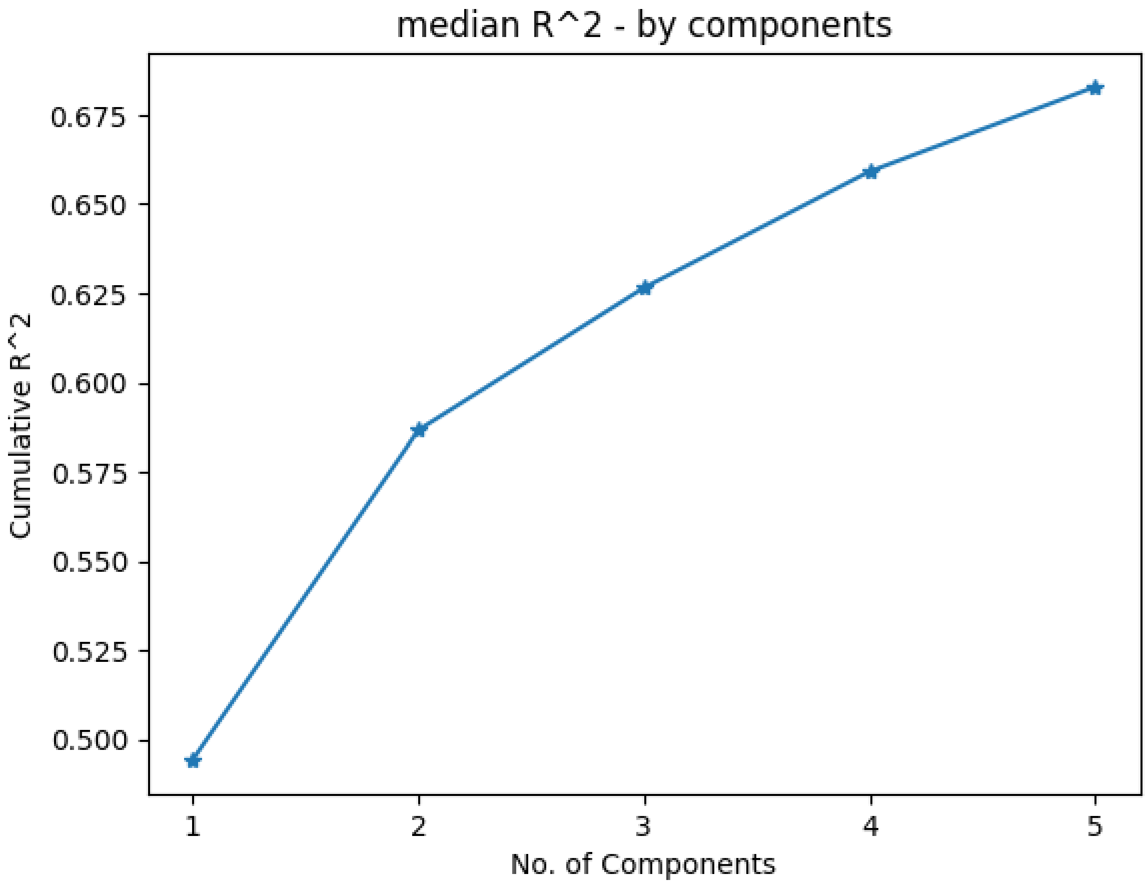


Figure 1 Median cumulative R^2 by components: for each running date, we calculate accumulative R^2 by components and plot the median values.

As shown in the chart, the R-squared values for the 2nd to 5th components are much smaller compared to the 1st component.

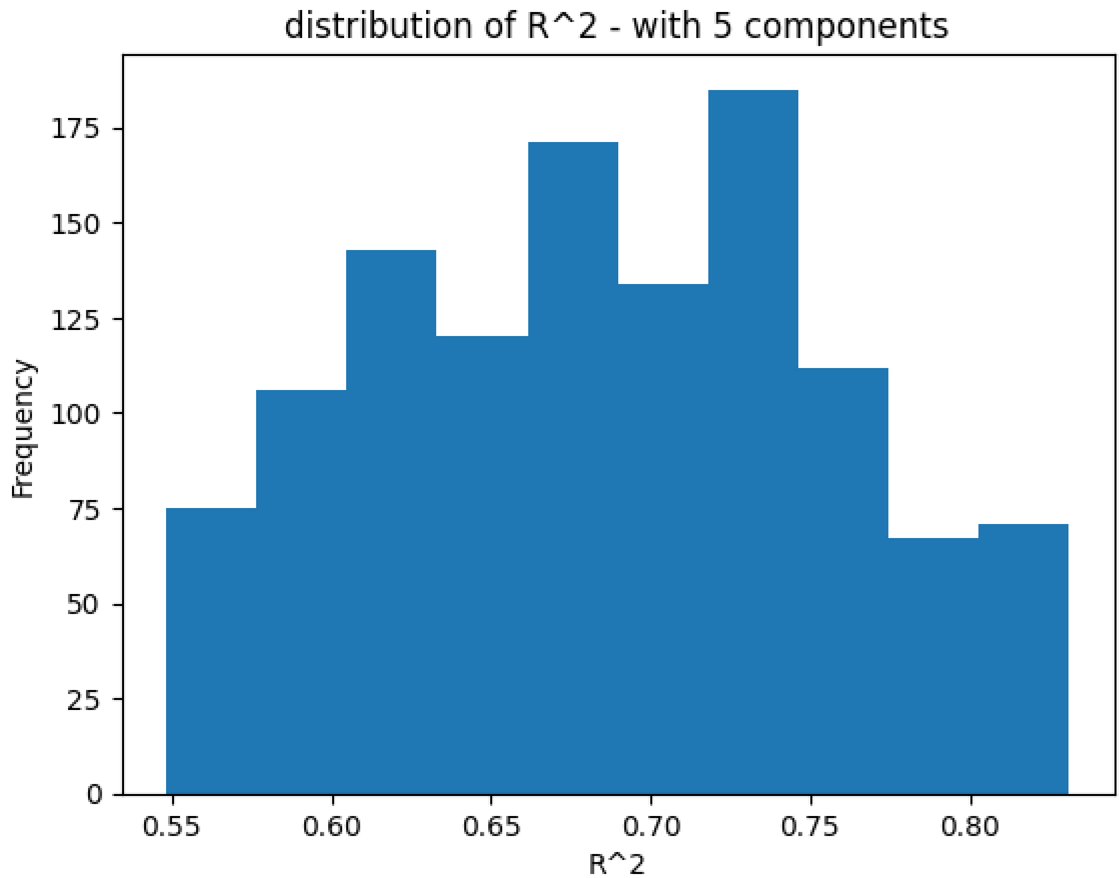


Figure 2 Distribution of cumulative R^2 with 5 principal components: we produce R^2 with all 5 components for each running date.

The histogram above shows that the R-squared values are relatively stable, with a minimum value around 0.55.

FIRST PRINCIPAL COMPONENT VS MARKET FACTOR

In general, principal components cannot be mapped to fundamental factors, except for the first one, which often corresponds to the general market factor. Before comparing these two models, it is interesting to examine the first principal component factor and market factor. It has been shown that these two are highly correlated (with a correlation greater than 0.97).

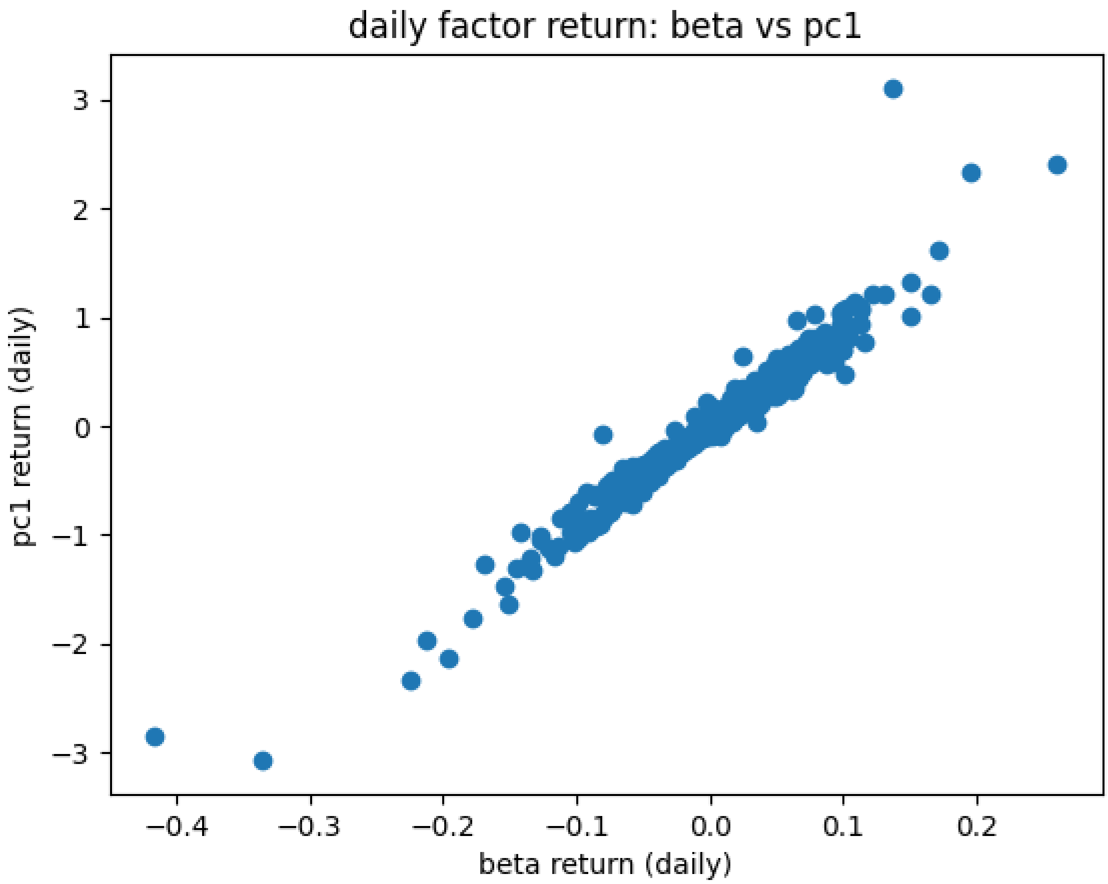


Figure 3 Comparison of daily beta factor return vs first principal component factor.

FACTOR VS PCA: IN-SAMPLE ANALYSIS

Next, we compare in-sample model R-squared, which shows how well the data fit the linear model. When calibrating factor exposures, we fit the asset return time series against factor returns for a specific time window. For each model run date we record R-squared (and p-value which are presented below) for each asset and plot the median value for that date, which is indicative of in-sample performance. It is noted that both models produce similar in sample R-squared level with PCA model slightly higher (by ~0.02).

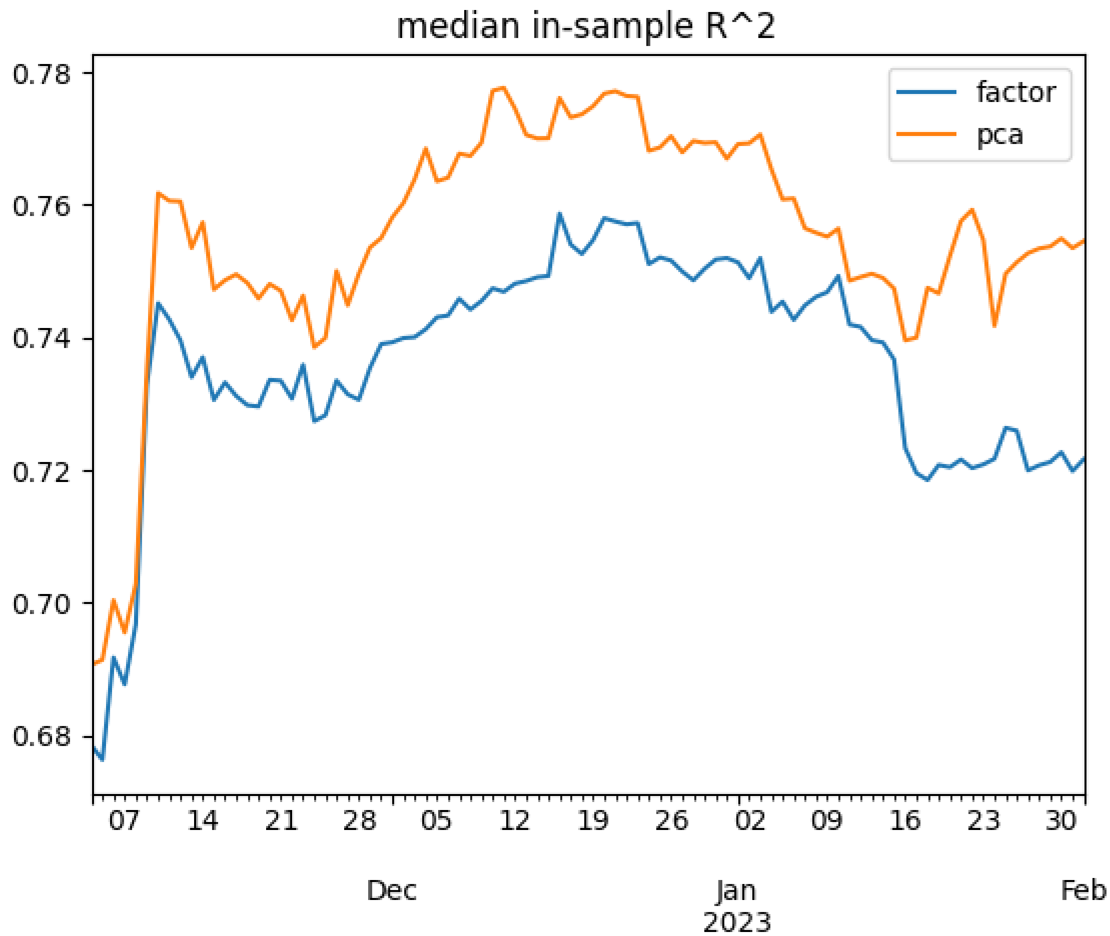


Figure 4 Median in-sample r-square. For each day, we run regression of asset return against factor returns for all assets within our universe, the plot shows the median r-squares for both factor model and PCA model.

We also list the median p-values of the factors and principal components side by side. It is seen that in-sample significance level for principal components are in general higher compared to SFRM factors.

factor	median p-value	principal component	median p-value
beta	0.000	pc1	0.000
size	0.192	pc2	0.031
liquidity	0.315	pc3	0.074
momentum	0.286	pc4	0.159
volatility	0.192	pc5	0.207

This is one of the metrics that we use to help improving our factor selection and consolidation, which is our current focus. The results in this table show that we still have some work to do in this area.

FACTOR VS PCA: OUT-OF-SAMPLE ANALYSIS

To find the model's out-of-sample explanatory power and significance level of each factor, we run weighted cross-sectional regression of forward asset return against asset factor exposure. For each run date, we check R-squared of the regression and p-value for each factor, R-squared represents the explanatory power of the risk models and p-values indicates significance level of corresponding factor. For both horizons that we tested, PCA model has slightly higher R^2.

	7-day horizon	30-day horizon
SFRM	0.289	0.309
PCA	0.298	0.318

The following two histograms show the distribution of r-square for both models.

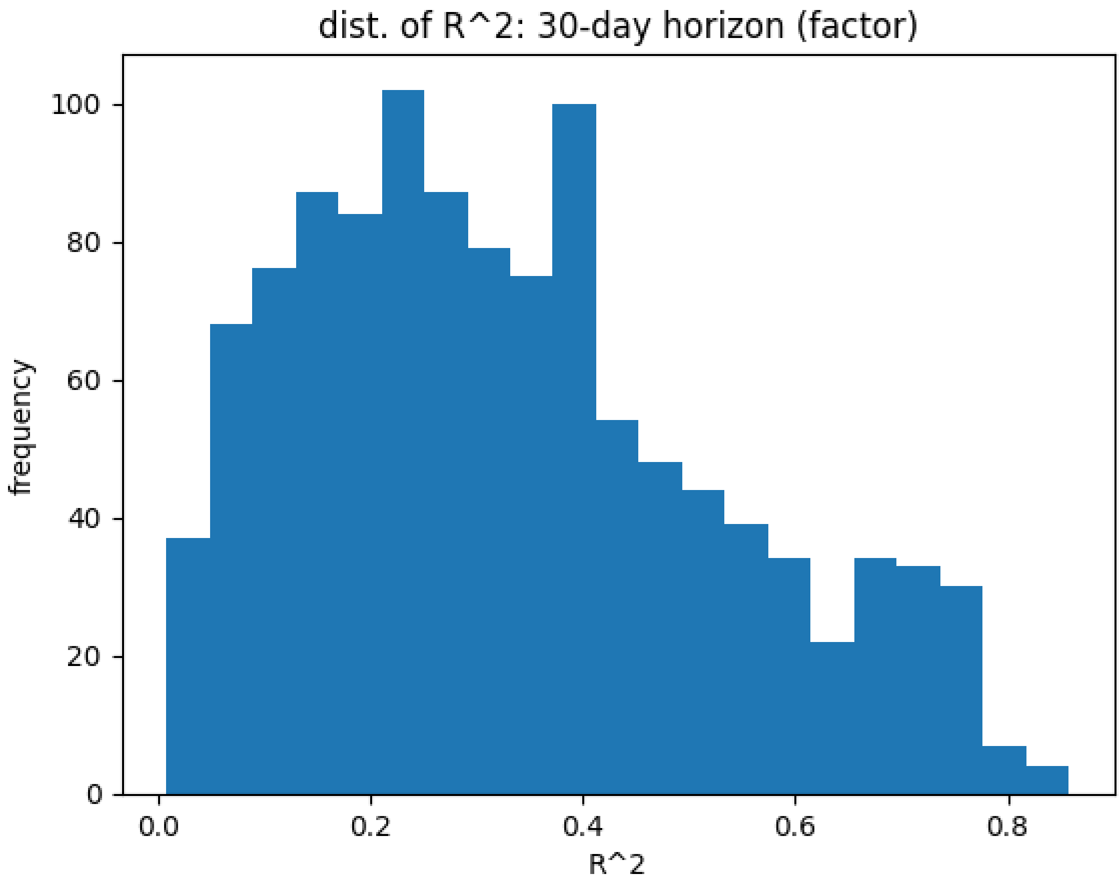


Figure 5 Distribution of out-of-sample cross-sectional R-squared for factor model.

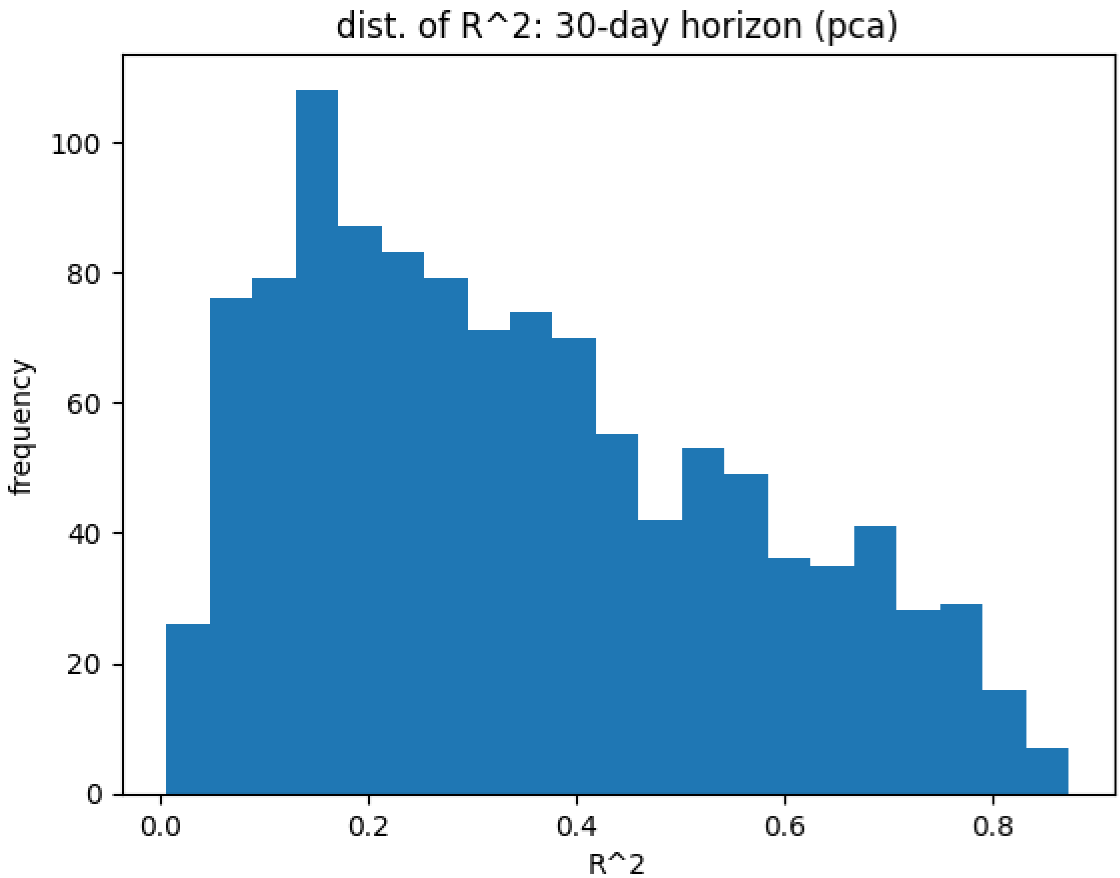


Figure 6 Distribution of out-of-sample cross-sectional R-squared for PCA model.

As for out-of-sample p-values, we present median values for 7/30-day horizon in the next two tables. For both horizons, SFRM factors are slightly more significant compared to PCA components. It is also worth noting that principal components for in-sample are more significant compared to SFRM factors.

SFRM	7-day	30-day	PCA	7-day	30-day
beta	0.01	0.00	pc1	0.00	0.00
volatility	0.09	0.13	pc2	0.22	0.28
size	0.15	0.12	pc3	0.20	0.21
liquidity	0.18	0.15	pc4	0.25	0.22
momentum	0.18	0.19	pc5	0.23	0.26

Again, we will keep on working on factors to improve the significance level.

FACTOR VS PCA: BIAS STATS

In addition to R-squared, bias statistic is another commonly used out-of-sample performance measure: the standard deviation of realized return to model predicted risk ratio. In an ideally scenario, this metric should be close to 1, however in reality this could not be exactly 1 given sampling error. To validate that our risk values are reasonable, we produce a list of portfolios and calculate the ratio for 30-day forward horizon: for each portfolio, we calculate the ratio of actual return and predicted risk. Time series for both models are plotted in the same graph, and they are very close. Please note that the ratios could spike due to high volatility at times, so the ratios are capped to remove outliers.

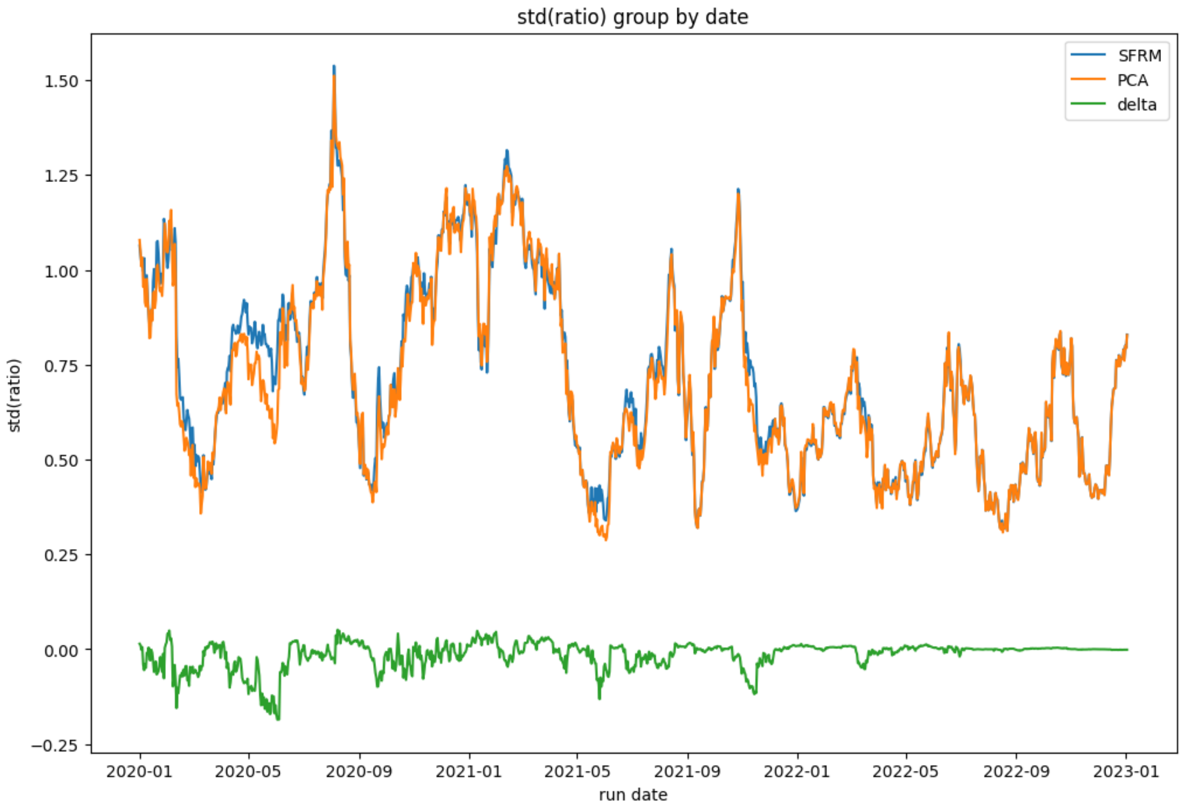


Figure 7 Plot of standard deviation of the ratios (portfolio return divided by predicted risk) for each running date.

CONCLUSIONS

In this report, we compare two approaches to model risks of digital assets: fundamental factor model vs PCA approach. In-sample as well as out-of-sample performance are compared: two approaches produce similar R-squared both in-sample and out-of-sample, bias stats are also quite close. While we continue improving these models, out-of-sample R-squared as well as bias stats metrics show that both models perform reasonably well, and it should be very helpful for the risk managing of crypto assets, especially in times when the overall market is very volatile.

It cannot be overstated that constructing a sophisticated risk model for digital assets is a laborious undertaking. While this analysis is not exhaustive, we have a lengthy list of upgrades to implement. Nevertheless, we wish to share our findings and receive input from the community. If you have any comments or feedback, please let us know.

Author

Chunlei Xia, Senior Quantitative Researcher @ Cloudwall Capital

REFERENCES

[1] R Grinold, R Kahn, "Active Portfolio Management", The McGraw Hill Companies (2000)
[2] D Shen, A Urquhart, P Wang, "A Three-Factor Pricing Model for Cryptocurrencies", Finance Research Letters (2020)
[3] Y Liu, A Tsyvinski, X Wu, "Common risk factors in cryptocurrency", The Journal of Finance (2022)
[4] L Cong, G Karolyi, K Tang and W Zhao. "Value Premium, Network Adoption, and Factor Pricing of Crypto Assets." Working Paper (2022)
[5] G Benigno and C Rosa, "The Bitcoin–Macro Disconnect", Federal Reserve Bank of New York Staff Reports, no. 1052 (2023)

Disclaimer

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

talos.com