

In the last report, several alpha factors have been discussed and tested on cryptocurrency market. However, the test was not rigorous. A systematic mathematical procedure is necessary to determine: **1) Whether the factor is significant? 2) Whether the factor is parameter-sensitive? 3) Is factor A more significant than factor B? 4) Is factor A independent from factor B?** (In another word, whether factor A provides extra information when factor B is known).

The questions proposed above must be solved in a quantitative way, instead of a subjective comparison between different plots.

## 1. Whether the factor is significant?

An intuitive way to determine whether a factor has predictive power is to look at its cumulative IC plot and see whether it is monotonically increasing or decreasing. Such method is subjective and can be quite inaccurate. We need a more objective test procedure.

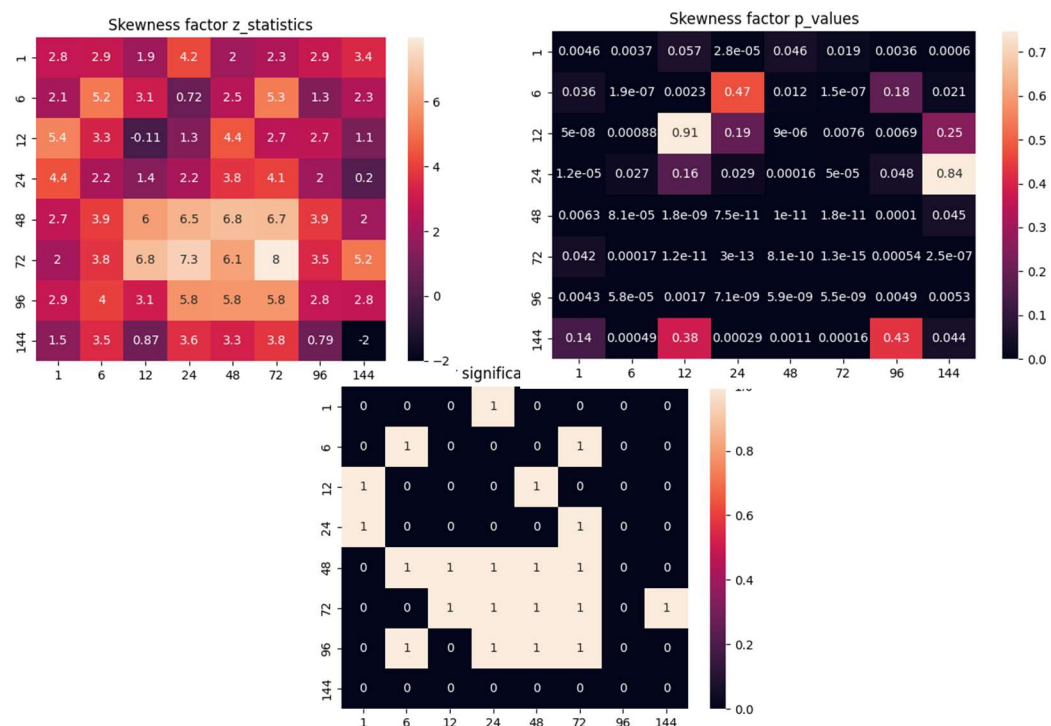
**Hypothesis Testing** is a good practice. If a factor indeed has predictive power on future price returns, its ICs should be significantly higher or lower than zero (comparing to random guess, whose IC should be averagely zero).

Due to the low signal-to-noise ratio in financial market, it is very likely that the outstanding predicting ability of a particular factor is actually a result of “luckiness”. To best avoid such false discovery, we set the p-value threshold to be extremely low. In the experiment below, only factors with p-value lower than 0.0001 will be concluded as significant predictors (Instead of conventional 0.05 or 0.01).

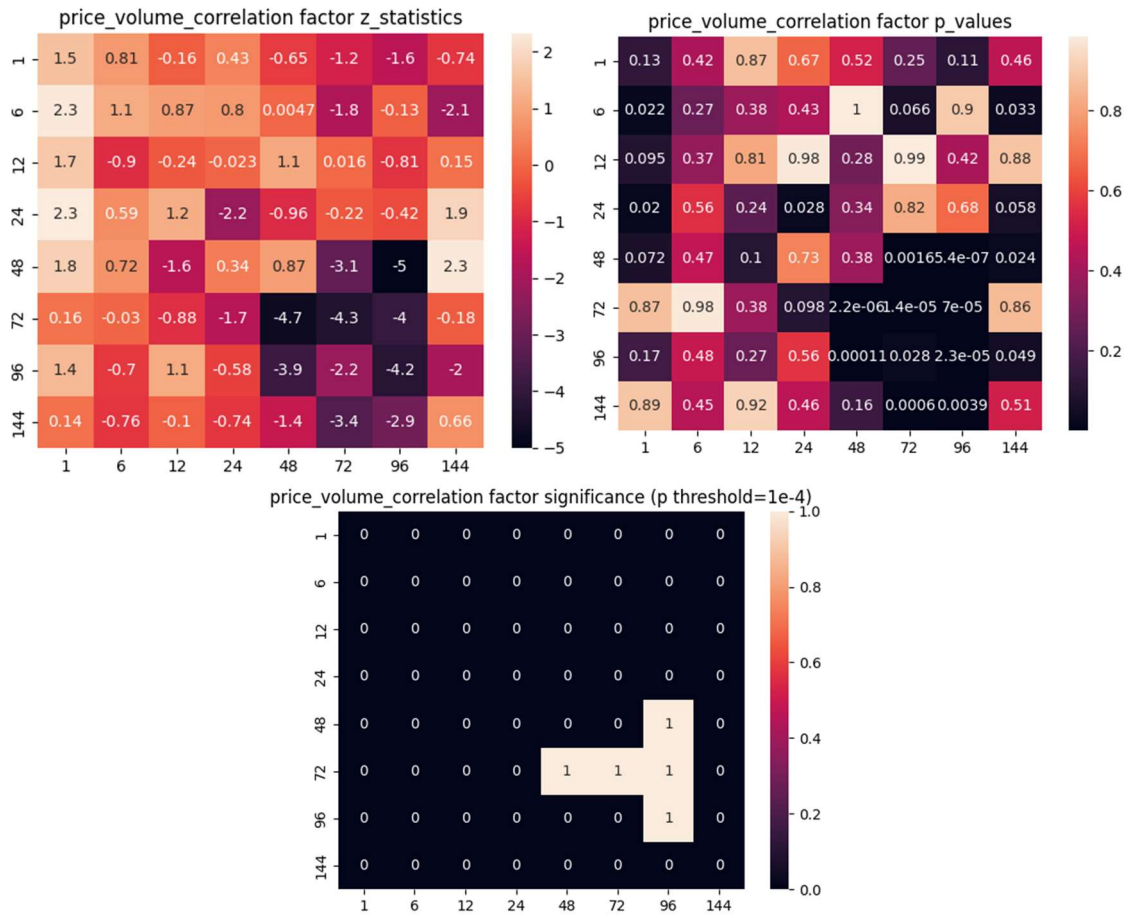
We perform z-test on the ICs of the 5 factors discussed in the last report.

*(In the figures below, x-axis is holding period, y-axis is lookback period)*

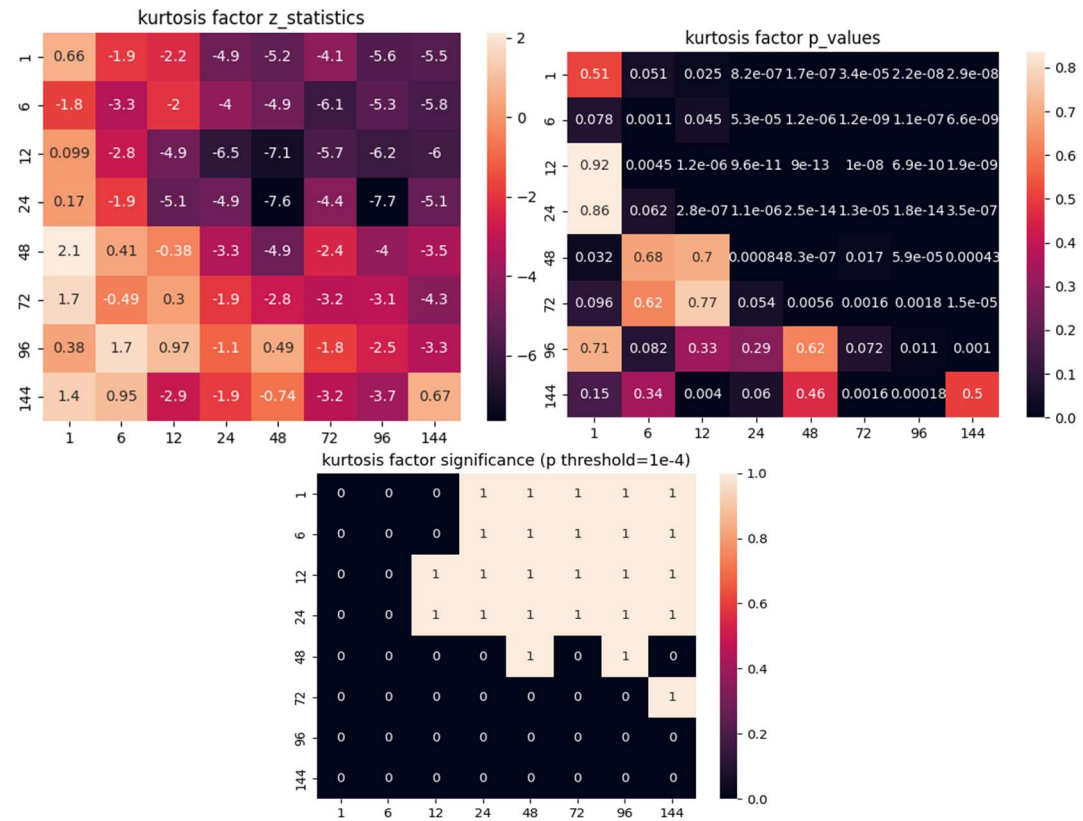
### 1) Skewness



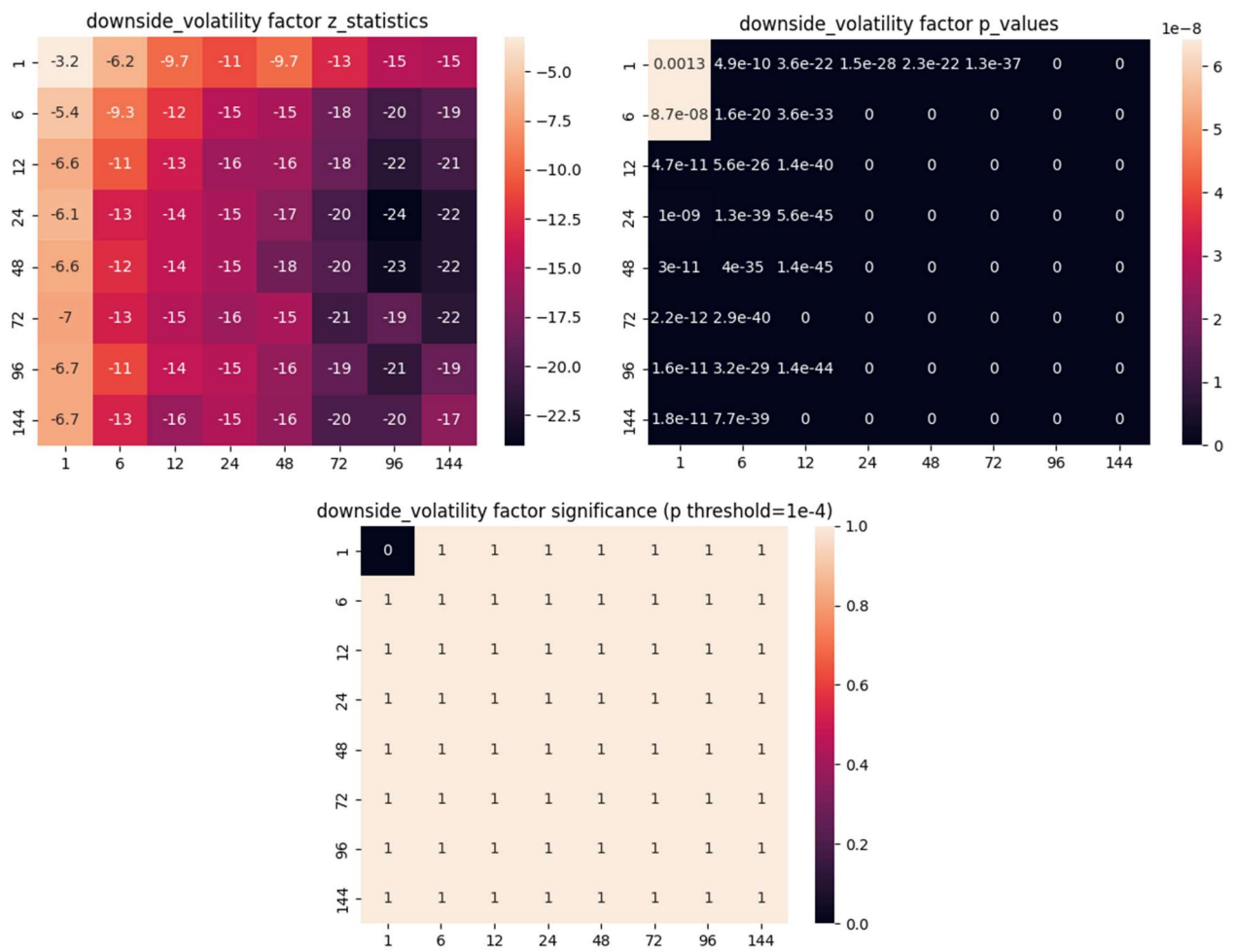
## 2) Price-Volume Correlation



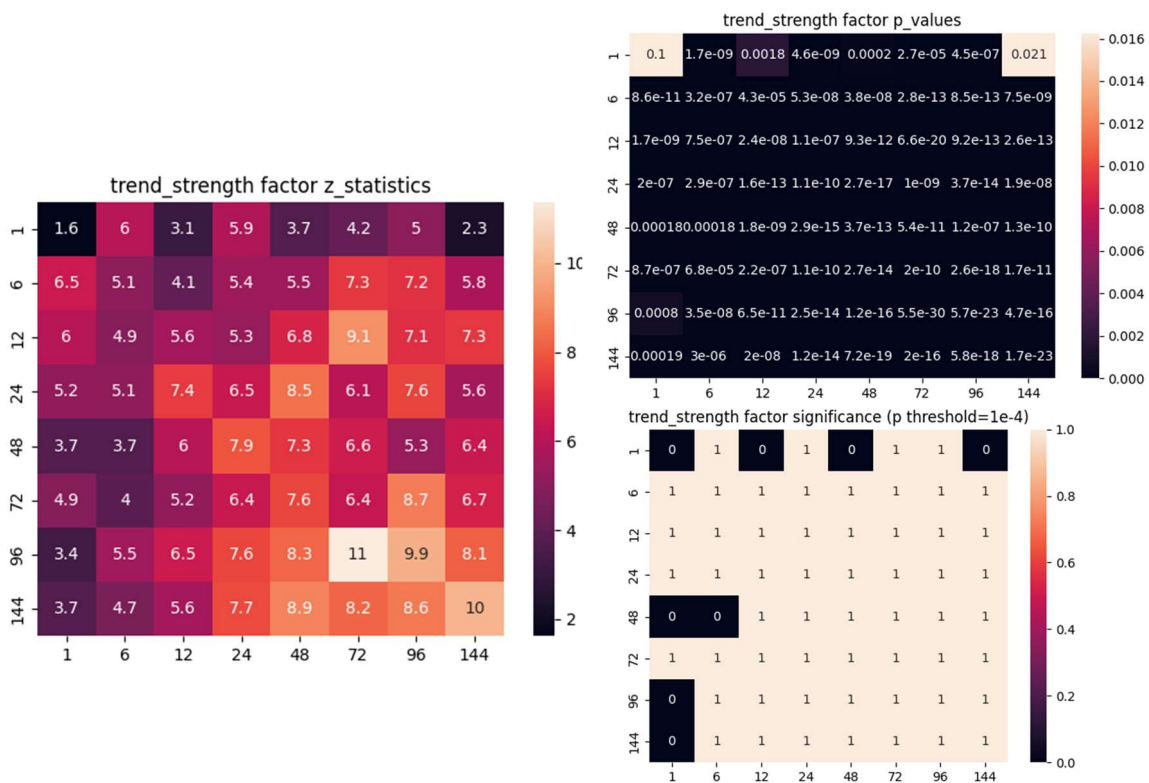
### 3) Kurtosis



#### 4) Downside Volatility



#### 5) Trend Strength



From the above figures, we find that all factors are significant within some parameter range, which will be further discussed in the next section.

## 2. Whether the factor is parameter-sensitive?

**Lookback** and **Holding** periods are essential parameters in every factor. Basically, Lookback period is how long of historical data we would like to use to calculate the latest factor value, and holding period is the time we wait for the next price observation (as the prediction result). Robust factors should not be parameter-sensitive. If a factor performs well under lookback=24, holding=24 (L24H24), then a nearby parameter set such (L25H23) should have similar performance. If the performance varies sharply as the parameters slightly change, it is very likely to be an overfit.

From the heatmaps above, we can observe that **downside volatility & trend strength** are significant under nearly all parameter combinations. As for **kurtosis & skewness**, their significant parameter sets cluster around some large area, which indicates they are robust to some extent. Although the significant parameter set of **price-volume correlation** also form a cluster, the area is relatively small so that we should be careful when we apply this factor.

## 3. Whether the factors are independent?

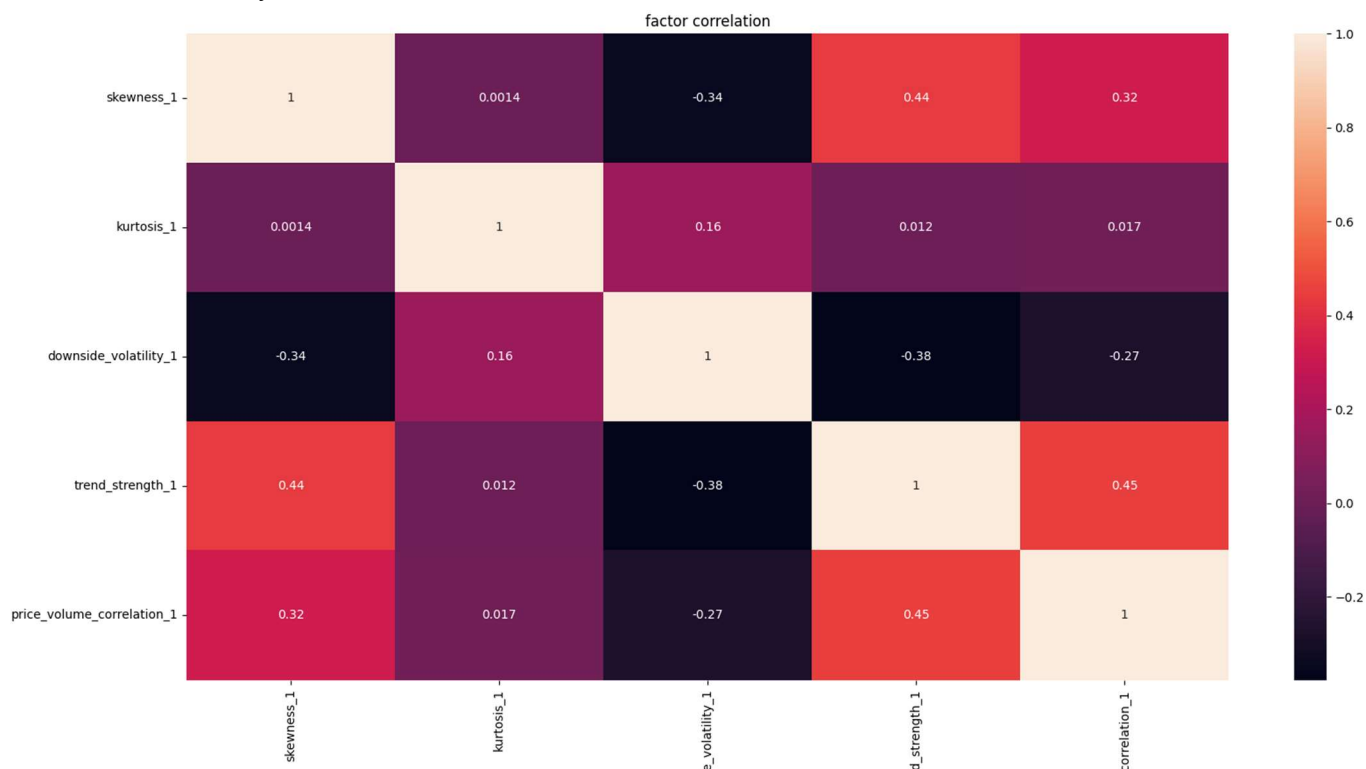
If we would like to utilize multiple factors for better prediction result than single factor, it is important that whether the factors are independent from each other. If one factor can be explained by the other, it cannot provide extra information for the prediction. Hence, we would like to find a set of factors that are not correlated, each of them contributes to the final prediction accuracy.

Intuitively, we may calculate the correlation matrix of the factors we are interested in. However, we cannot get a definite answer (yes or no) for whether a new factor can be used or not, only by looking at the correlation matrix. Obviously, we need new technique to help us accomplish this task

**Factor Orthogonalization**, widely used by both industry and academia, aims to exclude the influence of existing factors while testing a new factor. In nature, it takes the existing factor(s) as the explanatory variable(s) while the new factor as the explained variable, then performs a linear regression. The residuals are called **orthogonalized factor**, whose significance can be tested again. If the new factor can be represented by some linear combination of existing factors, these residuals should be white noise which contains no information. However, if there are unexplainable components in the residuals that still have predictive power, the new factor does provide extra information and is worth to be included.

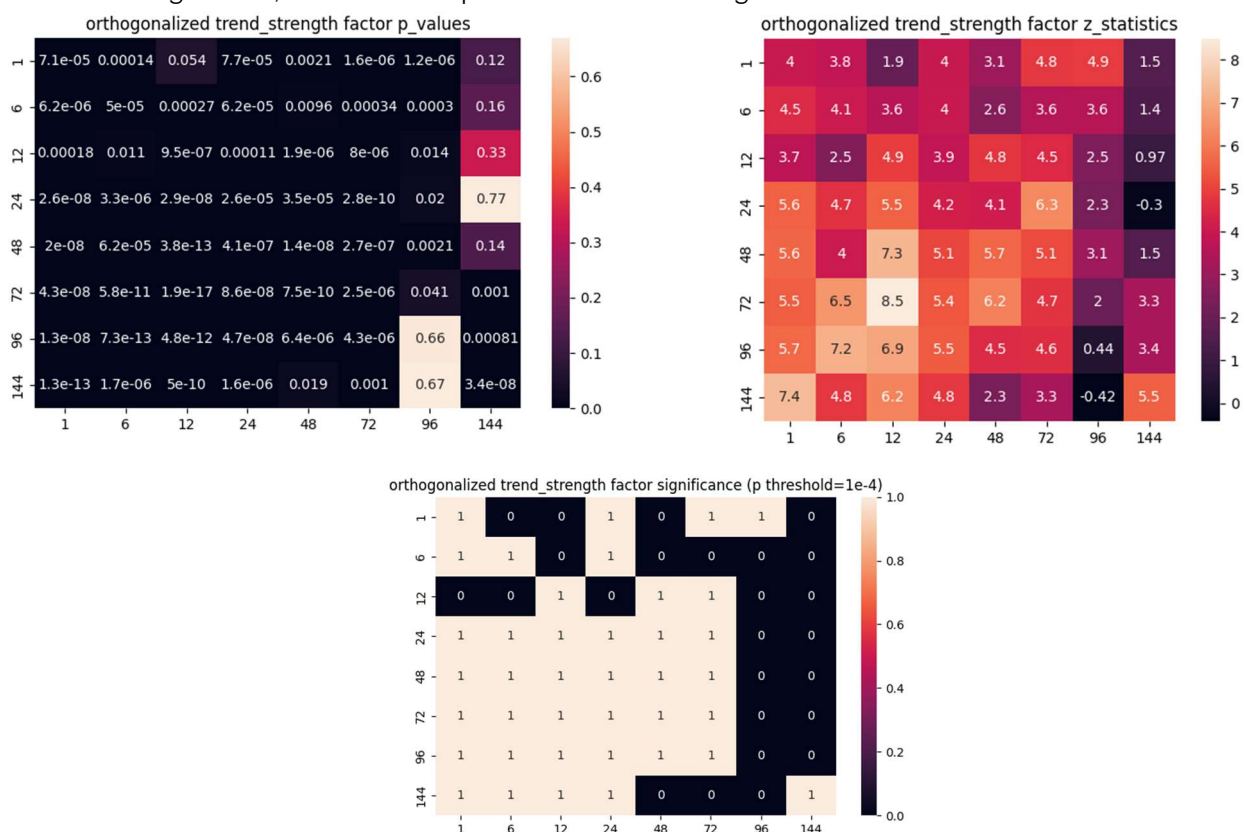
Since we have observed that **downside volatility** is the most significant factor, we will use it as the first known factor, and test the rest one by one.

Firstly, we calculate the correlation matrix of all factors, for reference.



As shown, all of the factors are not highly correlated with the others, which is a good indication. However, we still need to verify it by factor orthogonalization.

We use downside volatility as X variable while trend strength as y variable. After linear regression, the residuals' performance as orthogonalized factor is shown below.



This is a very simple trial on factor orthogonalization. It shows that trend strength provides extra information given downside volatility. However, this method only applies to two factors. If we would like to perform orthogonalization on multiple factors simultaneously, more complicated techniques are required, which will be left for the next report.

(Factor orthogonalization is a bit complicated, which needs further study, naive linear regression may not be accurate. Hence, this week ends here. Source code & figures are put in the folders anyway.)