URECA recurrent report – Sem 1, week 1~2

When the problem comes to predicting future returns of a particular commodity/security, the most important problem is to find out which features are useful for predictions, which is also called ***return predictors, or (alpha) factors.*** A good factor must be **robust**, **explainable** and **significant** enough. Hence, if one tries to dig or extract features through naïve machine learning (ML), it will end up **unexplainable (Blackbox)**. **Overfitting** is another serious problem that is very likely to encounter. This is because the amount of historical data of a particular commodity/security is far from enough for most ML algorithms. Therefore, to make a good start, a good choice is to use factors that has already been discovered, economically explainable, and has continuous good performance for a long time in the past. Feeding such features to the ML algorithms can be much better than feeding raw data.

This report has verified the effectiveness of several factors calculated based on 1-min price & volume information. Instead of Open, High, Low, Close, Volume (OHLCV) data, high frequency data is expected to contain more information. Limit order book data is another good choice for constructing factors, however it is very expensive to purchase, we will leave it for future research.

Factors that have been studied in this report including:

1. **Skewness**

   Skewness of price movements in a certain period (I.e., 1hour, 1day, etc.) is defined as:

$$Skewness = \frac{\sqrt{N}\ \sum_{j=1}^{N} r_j^3}{(\sum_{j=1}^{N} r^2)^{1.5}}$$

   Where j represents the jth minute of this period.

2. **Kurtosis**

   Kurtosis of price movements in a certain period is defined as:

$$Kurtosis = \frac{\sqrt{N}\ \sum_{j=1}^{N} r_j^4}{(\sum_{j=1}^{N} r^2)^{2}}$$

3. **Downside Volatility**

   Downside Volatility is the percentage of downside volatility that contributed to total volatility. Downside Volatility is defined as follows:

$$Downside\ Volatility = \frac{\sum_j^N r_j^2 I_{r_j<0}}{\sum_j^N r_j^2}$$

4. **Price-Volume Correlation**

   The correlation coefficient between intraday price and volume, varies from 0 to 1.

$$Price - Volume\ Correlation = \ corr(Close_j, \frac{Vol_j}{\sum_j Vol_j})$$

5. **Trend Strength**

   Trend Strength is the ratio between the price movement distance and absolute length of path.

$$Trend\ Strength = \frac{P_n - P_1}{\sum_{j=2}^{N} |P_j - P_{j-1}|}$$

   Where P_j represents the price at time j.

How to test factor effectiveness?

Factor effectiveness can be measured in the form of ***Information Coefficient (IC).*** This statistic measures the correlation between the latest factor value and future return within some time. A common practice is calculating ***RankIC,*** which is the correlation coefficient between the rankings of latest factor value and rankings of future return. If the RankIC of a particular factor is significantly greater than zero or less than zero, we say this factor has predictive power.

This experiment is conducted on 18 most liquid cryptocurrencies. Unlike traditional commodity futures, cryptocurrencies are traded in 7x24 scheme, which means the data is more sufficient, continuous and easy-to-clean.
(I also tried on commodity futures but there seems some data quality related problem. Should find some more reliable free data source in the future)

The rest of the report will present IC matrices of the factors mentioned above. IC matrices contains RankIC of different review period and holding period. The period unit is chosen to be 1 hour instead of 1 day. This is because the market is usually more predictable under high-frequency time frame empirically.
(The p-statistic test is not done due to time limit, will be done in the future. Currently, the significance can be roughly judged by printing out the cumulative IC plot.)

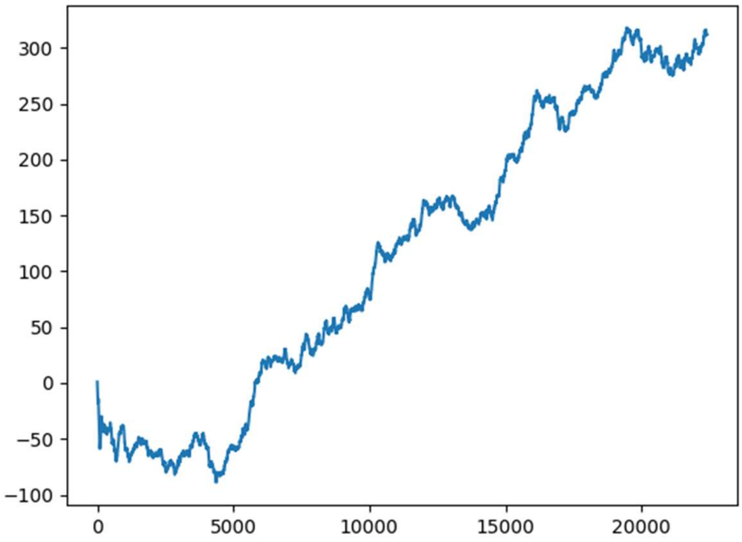
1. Skewness

Hold

Lookback

Index	1	6	12	24	48	72	96	144
1	0.00485918	0.00502006	0.00330936	0.00718831	0.0034432	0.00403909	0.00501583	0.00593239
6	0.00361744	0.00883695	0.0052194	0.00123389	0.00434324	0.00894782	0.00227686	0.00396455
12	0.00932958	0.00566025	-0.000187269	0.00224922	0.00761745	0.0045684	0.00469339	0.00196576
24	0.007499	0.00380357	0.00240577	0.0037683	0.00651447	0.00706315	0.00344104	0.00034229
48	0.00468149	0.00679297	0.0104099	0.0111903	0.0118485	0.011632	0.00680115	0.00353317
72	0.00347986	0.00639757	0.0117012	0.0126267	0.010638	0.0138428	0.00593777	0.00890072
96	0.00485715	0.00679937	0.00539765	0.0100365	0.0100771	0.0101177	0.00482011	0.00477492
144	0.00252044	0.00590491	0.00146898	0.0061619	0.00551758	0.00643919	0.00133435	-0.00344893

Cumulative IC plot of Hold=72, Lookback=72:



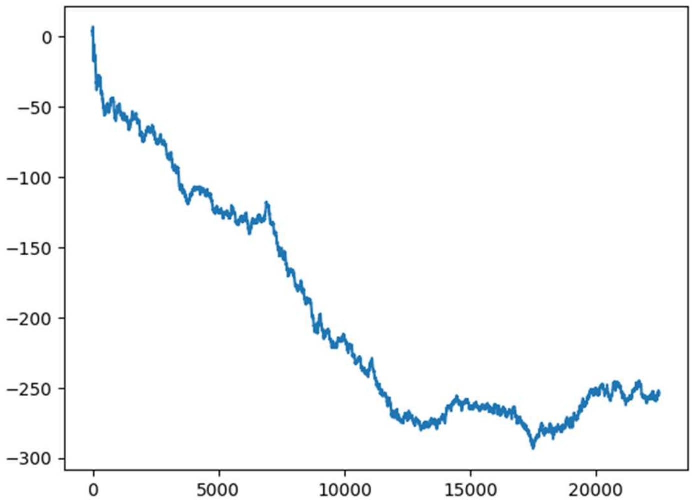
2. Kurtosis

Hold

Lookback

Index	1	6	12	24	48	72	96	144
1	0.00112301	-0.00335071	-0.00386831	-0.0085208	-0.00901003	-0.00709646	-0.00968918	-0.00974187
6	-0.00304472	-0.00557827	-0.00347175	-0.00694731	-0.0084113	-0.0105837	-0.00914926	-0.0101293
12	0.000171235	-0.00493914	-0.00830545	-0.0112053	-0.0123319	-0.00993116	-0.0107118	-0.0105514
24	0.000298202	-0.00319198	-0.00880782	-0.00845046	-0.013203	-0.00750123	-0.0133464	-0.00895124
48	0.00363514	0.000696618	-0.000657591	-0.00580745	-0.00841452	-0.00413015	-0.00694262	-0.00611264
72	0.00282911	-0.000836125	0.000513768	-0.00327343	-0.00479315	-0.00550058	-0.00544821	-0.00756173
96	0.000643719	0.00294198	0.00166646	-0.00184311	0.000855623	-0.0030981	-0.00438384	-0.0057425
144	0.00242364	0.00161856	-0.00491903	-0.00319472	-0.00126927	-0.00541098	-0.00638829	0.00117099

Cumulative IC plot of Hold=24, Lookback=12



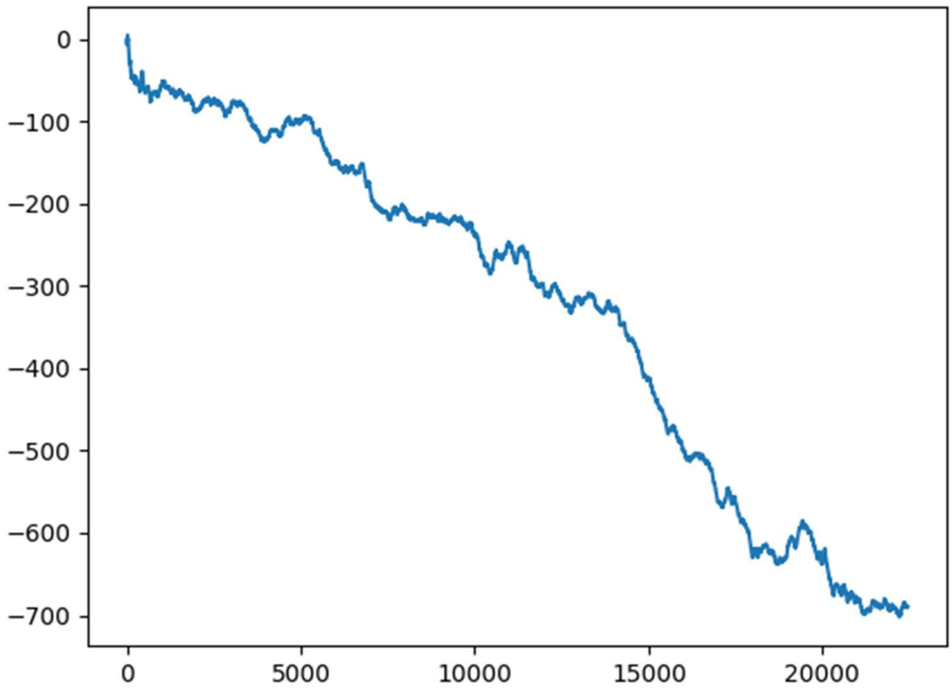
3. Downside Volatility (Very Significant!)

Hold

Lookback

Index	1	6	12	24	48	72	96	144
1	-0.00560364	-0.0107282	-0.0167071	-0.0191103	-0.0166165	-0.0221887	-0.0251031	-0.0262209
6	-0.00933961	-0.0162067	-0.0207677	-0.0252912	-0.0265114	-0.0316631	-0.0350159	-0.0338841
12	-0.0114796	-0.0182839	-0.0231313	-0.027321	-0.0276374	-0.0318675	-0.0373392	-0.0358226
24	-0.0106645	-0.0229416	-0.0243568	-0.0262803	-0.0288707	-0.0346537	-0.0413501	-0.0388808
48	-0.011558	-0.0212925	-0.0244106	-0.0261766	-0.0306018	-0.0336494	-0.0401987	-0.0386098
72	-0.0120288	-0.0226707	-0.0252938	-0.0271379	-0.0260291	-0.0353343	-0.0326383	-0.0379007
96	-0.0115069	-0.0190937	-0.0239362	-0.0250158	-0.0276979	-0.0322388	-0.036699	-0.0319756
144	-0.0114265	-0.0221375	-0.0272245	-0.0253025	-0.0278019	-0.033881	-0.0337778	-0.0284645

Cumulative IC plot of Hold=24, Lookback=24



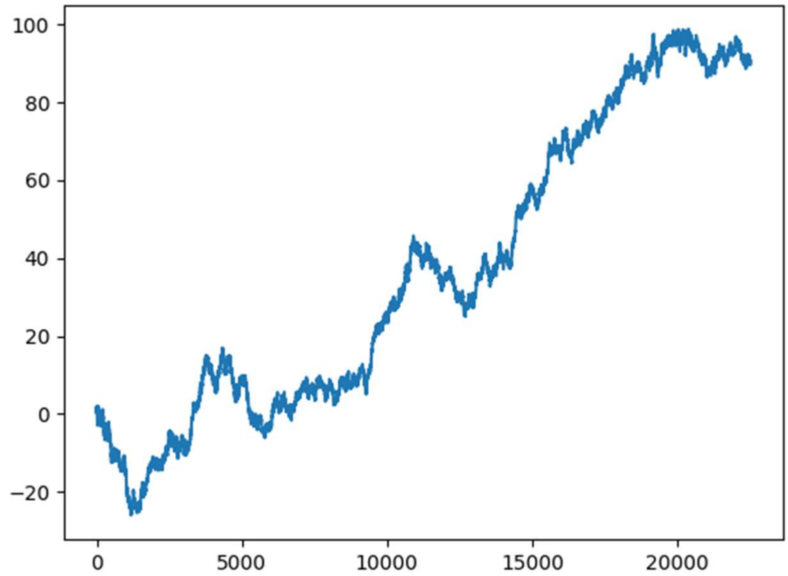
4. Price-Volume Correlation (Not very good...)

Hold

Lookback

Index	1	6	12	24	48	72	96	144
1	0.00261184	0.00139612	-0.00027631	0.000744513	-0.00111612	-0.00199039	-0.00272189	-0.00127714
6	0.00394321	0.00190923	0.00149759	0.00136694	8.04449e-06	-0.0031792	-0.000217826	-0.0037014
12	0.00289053	-0.00156775	-0.000405904	-4.03874e-05	0.00186373	2.77602e-05	-0.00140255	0.000253259
24	0.00400825	0.00101482	0.00204333	-0.00375888	-0.00166214	-0.000381756	-0.000719763	0.00330065
48	0.00309831	0.00122083	-0.00280637	0.00059024	0.00148749	-0.005411	-0.00858126	0.00391928
72	0.000275111	-5.17529e-05	-0.00150715	-0.00282324	-0.00810094	-0.00747997	-0.00687635	-0.000305167
96	0.00236749	-0.00118672	0.00188778	-0.000987205	-0.00658091	-0.00378309	-0.00735472	-0.00341823
144	0.00024033	-0.0012916	-0.000168216	-0.00125269	-0.00238279	-0.00586582	-0.00493776	0.00113321

Cumulative IC plot of Hold=1, Lookback=24



5. Trend Strength

Hold

Lookback

Index	1	6	12	24	48	72	96	144
1	0.00279598	0.0103734	0.00540047	0.0101333	0.00640716	0.007209	0.00867711	0.00397348
6	0.0112259	0.00889931	0.00707853	0.00950584	0.00946673	0.0125111	0.0123314	0.00996426
12	0.0103958	0.00857141	0.00957801	0.00926167	0.0117497	0.01579	0.0123043	0.012638
24	0.0089833	0.00883027	0.0127513	0.0111702	0.0146535	0.0104711	0.012924	0.00972772
48	0.00641764	0.00643022	0.0103216	0.0134863	0.012384	0.0111847	0.00908636	0.0110254
72	0.00842457	0.00685533	0.00884781	0.0110299	0.0129297	0.010889	0.0149269	0.0115668
96	0.00568914	0.00939587	0.0111196	0.0128427	0.0139631	0.0194663	0.0169172	0.014013
144	0.00634574	0.00794394	0.00955248	0.0131294	0.015143	0.0141705	0.0148532	0.0172068

Cumulative IC plot of Hold=24, Lookback=24