1. Executive Summary

Factor model is spanned across the whole life cycle of quantitative equity world, for a variety of investors from the entire spectrum of the active risk budget. There are two main streams of factor model estimation from the highest level:

(1) In the academia, researchers usually prefer time series regression (a.k.a spanning test) to validate the superiority of their new factor to explain the market anomaly beyond traditional 3/4/5/6 factor framework. They start with a target return series and a set of pricing factor series (say k factor, where each factor is a zero-cost L/S portfolio), and get k scalars as the factor loadings. Those zero-cost L/S portfolios could be considered as the difference between two unlevered rates of return between long leg and short leg, coming with some issues, like we cannot guarantee the exposure to the certain characteristic is always 1 or the exposure to other characteristics is always 0.

(2) In the industry, we usually consider the investment universe cross-sectionally in both the risk and alpha frameworks, but the nuance in estimating and evaluating various types of signal differs. In the risk model, a majority of asset managers follow the principle of "return neutralization": in each cross-section, factor exposures are fixed while factor returns are estimated through regression. The return to a factor is sensitive to the set of other factors included in the model, while the exposure is fixed. The portfolio's exposure is essentially the weighted sum product. In the alpha model, one could still pursue the academic-style regression to showcase her "skill" to beat the market as a smart active investor, but most of time, the result is super time-varying and hard to be further attributed to the downstream (e.g. tilt v.s. timing). Hence, residing in the 80's/90's bay area, Prof. R Grinold developed the holding neutralization framework to ingest fund-specific holding information into the regression. In this setup, return to a given source is fixed but the exposure is estimated through regression.

Section 2 goes through the programming process with some highlighted results. Section 3 shares some insights on signal neutralization and answers the open questions. Section 4 briefly describes the quality control. Section 5 concludes and discusses potential extension work.

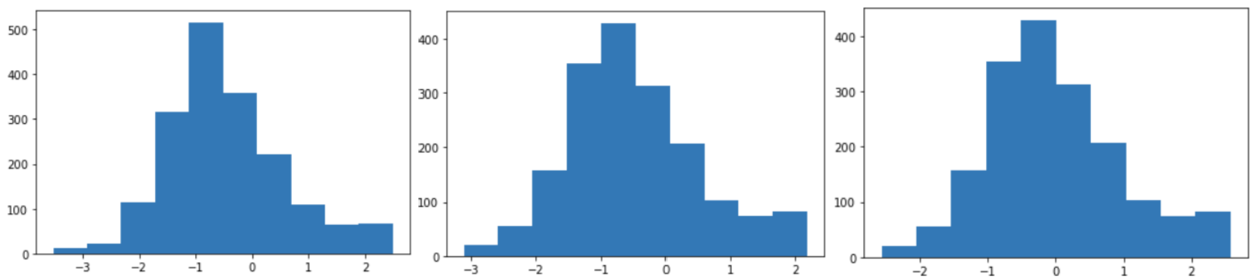
2. Programming and Result

Following the instruction, I ran the cross-sectional analysis using the combined universe to get 4 different factor returns matrices based on 4 approaches (naïve OLS, see the file *"facRets_ols"*; naïve WLS, see the file *"facRets_wls"*; optimization-based WLS with industry return constraint, see the file *"facRets_wls_constrain"*; L/S zero-cost portfolio, see the file *"facRets_LS"*). After this, the factor returns are assumed to be fixed in the testing stage.

In the data exploration stage, I first check the data quality and dimension. In the cross-section, I filled exposure data based on market cap and industry, with outlier removal, standardization and etc. In the time series, I forward filled missing data in time series to pursue a longer history as much as possible.

I am grateful for valuable insights gained from countless informal communication with quant equity practitioners from academia and industry, either within the organization (Dr. V Chandrashekaran) or beyond (Dr. W Choi), to name a few.

The below histogram demonstrates the raw CNTRD_MOMENTUM distribution on Nov 8, 2021, with MAD-based outlier removal, then standardization, from left to right.



On the universe selection, since index constitution is updated semiannually, we used the Dec 2020 and Jun 2021 rebalanced universe as the most up-to-date ones. Specifically, Jun 15, 2021 is the first day that we use new universe returns, so Jun 14, 2021 is the first day that using new universe exposures, which means Jun 13, 2021 is the latest possible date we use old universe exposure.

For OLS and WLS estimation, I leveraged the *statsmodels* package to run the analysis in each cross-section, where the only difference is WLS use the square root of market cap as weighting matrix. The factor scope includes country, industries, styles.

The devil in the detail is that the above one-pass cross-sectional regression doesn't constrain the sum of industry-level returns to be zero, even though the resulted factors returns have been neutralized. For each asset, there exist two intercept terms – industry and country. Every asset has unit exposure to them. This can be resolved by adding a linear restriction on the factor returns. Intuitively, it represents a rotation of the factor returns, while doesn't impact the efficacy of the risk model. To illustrate, if asset returns across the country universe are mostly positive on a given day and healthcare are also up but by less than the average, then the healthcare factor return will be negative.

I used stock-level square root of market cap to get the industry-level weight to keep it in an affine form and implemented through *scipy.optimize* package. From the figure below, the country and industry factor returns by WLS_constrain differs from WLS, while the style factor returns are not touched.

In [202]:
```
df_facRets_wls_constain.T[Styles].head()
```
Out[202]:

	CNTRD_ANLYSTSN	CNTRD_BETA	CNTRD_BTOP	CNTRD_DIVYILD	CNTRD_EARNQLTY	CNTRD_EARNVAR	CNTRD_EARNYILD	CNTRD_GROWTH
2021-11-01	-0.001079	0.000624	-0.002536	-0.000652	0.000951	-0.001333	0.004870	-0.003128
2021-11-02	0.001040	-0.003478	-0.001408	-0.001011	-0.000643	-0.002461	-0.004345	0.002108
2021-11-03	-0.000289	0.001164	-0.000106	0.000845	-0.000858	0.001094	0.003830	0.000353
2021-11-04	-0.000570	-0.002209	-0.002547	0.000882	-0.001933	-0.001840	-0.004090	0.000660
2021-11-05	-0.000265	-0.000790	0.000447	0.000048	-0.001095	-0.001073	-0.002629	0.001137

In [201]:
```
df_facRets_wls.T[Styles].head()
```
Out[201]:

	CNTRD_ANLYSTSN	CNTRD_BETA	CNTRD_BTOP	CNTRD_DIVYILD	CNTRD_EARNQLTY	CNTRD_EARNVAR	CNTRD_EARNYILD	CNTRD_GROWTH
2021-11-01	-0.001079	0.000624	-0.002536	-0.000652	0.000951	-0.001333	0.004870	-0.003128
2021-11-02	0.001040	-0.003478	-0.001408	-0.001011	-0.000643	-0.002461	-0.004345	0.002108
2021-11-03	-0.000292	0.001174	-0.000106	0.000846	-0.000855	0.001099	0.003831	0.000353
2021-11-04	-0.000570	-0.002209	-0.002547	0.000882	-0.001933	-0.001840	-0.004090	0.000660
2021-11-05	-0.000265	-0.000790	0.000446	0.000049	-0.001095	-0.001073	-0.002629	0.001137

I am grateful for valuable insights gained from countless informal communication with quant equity practitioners from academia and industry, either within the organization (Dr. V Chandrashekaran) or beyond (Dr. W Choi), to name a few.

In [209]:
```
df_facRets_wls_constain.T[['CNTRD_COUNTRY']+Industries].head()
```
Out[209]:

	CNTRD_COUNTRY	CNTRD_AEROSPCE	CNTRD_AIRLINES	CNTRD_APPAREL	CNTRD_AUTOCOMP	CNTRD_BANKS	CNTRD_BEVTOB	CNTRD_BLDPR(
2021-11-01	0.011850	0.016825	-0.008834	-0.017832	-0.004627	0.014392	-0.005840	-0.0093
2021-11-02	-0.010805	0.033277	0.011517	0.003482	0.018967	-0.010790	-0.006569	0.0012
2021-11-03	0.006286	-0.014374	0.001020	0.010445	-0.008815	-0.006553	-0.003970	-0.0051
2021-11-04	0.011430	-0.000820	-0.008433	-0.008667	0.023816	-0.004633	0.019216	-0.0113
2021-11-05	-0.007716	-0.007996	-0.002278	0.011311	0.016112	0.005864	0.005654	-0.0092

5 rows × 33 columns

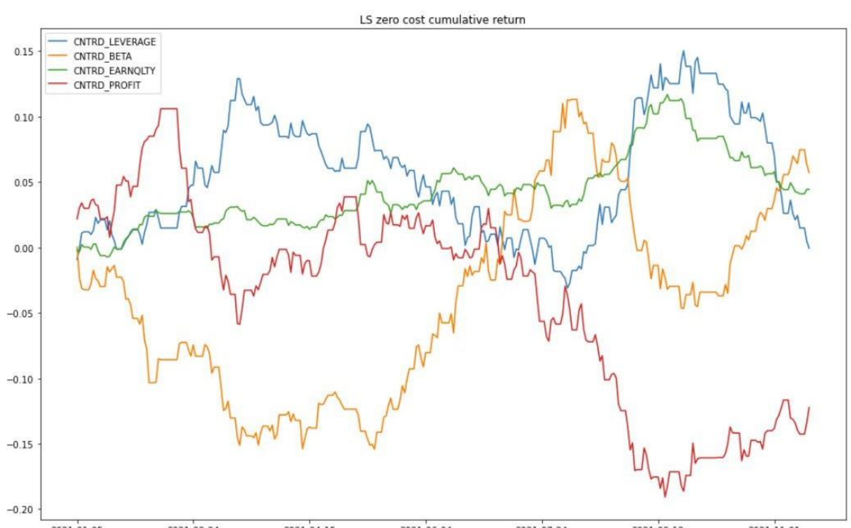
In [210]:
```
df_facRets_wls.T[['CNTRD_COUNTRY']+Industries].head()
```
Out[210]:

	CNTRD_COUNTRY	CNTRD_AEROSPCE	CNTRD_AIRLINES	CNTRD_APPAREL	CNTRD_AUTOCOMP	CNTRD_BANKS	CNTRD_BEVTOB	CNTRD_BLDPR(
2021-11-01	0.007372	0.021303	-0.004356	-0.013354	-0.000149	0.018870	-0.001362	-0.0049
2021-11-02	-0.010574	0.033046	0.011287	0.003251	0.018736	-0.011021	-0.006800	0.0010
2021-11-03	0.006836	-0.014883	0.000463	0.009809	-0.009380	-0.007105	-0.004554	-0.0056
2021-11-04	0.010228	0.000381	-0.007230	-0.007464	0.025018	-0.003431	0.020418	-0.0101
2021-11-05	-0.009223	-0.006474	-0.000750	0.012829	0.017619	0.007376	0.007167	-0.0077

5 rows × 33 columns

The "Long Top 20%/Short Bottom 20%" is self-explanatory. I use equal-weight for both legs in this exercise to keep more volatility characteristic using those small-cap names. The factor scope includes styles.

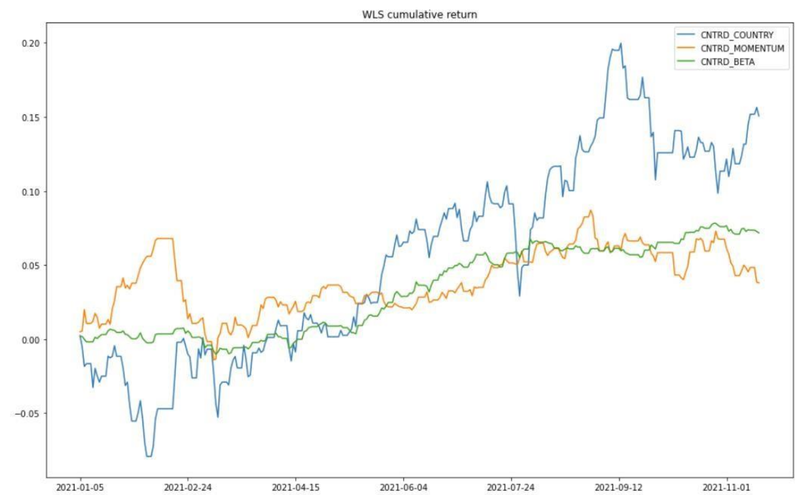
In the testing stage, firstly, I calculated and plotted the cumulative returns of holding various risk FP in the combined universe based on the 4 FP construction approaches. Again, *WLS_constrain* is the recommended.



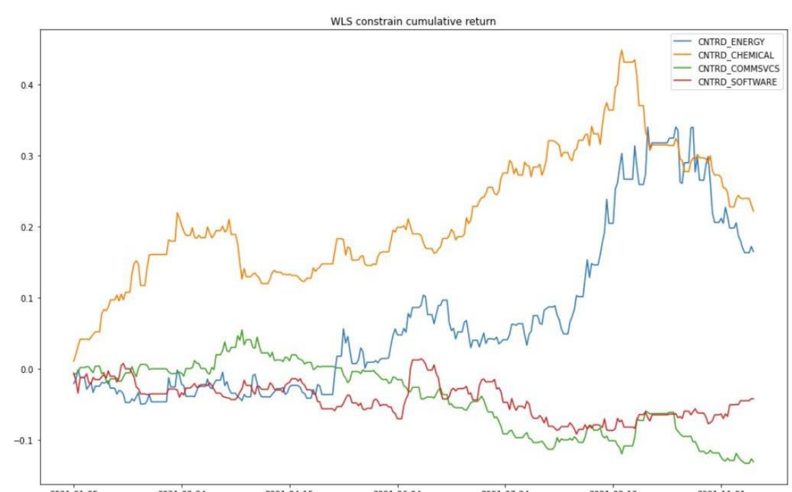
One take from this chart is "fly to quality": when the L/S beta portfolio relatively underperformed, the high quality names may outperform, depending on the overall market risk appetite. By the way, if the rocketing bitcoin price during the Mar 2023 banking industry crisis does not make sense to market participants, it is also partially due to "fly to quality", on a relative basis.

I am grateful for valuable insights gained from countless informal communication with quant equity practitioners from academia and industry, either within the organization (Dr. V Chandrashekaran) or beyond (Dr. W Choi), to name a few.

From the above, another point I want to make is the negative skew of the momentum factor return. The crash is always more severe around the bear-to-bull turning point than the other way around.



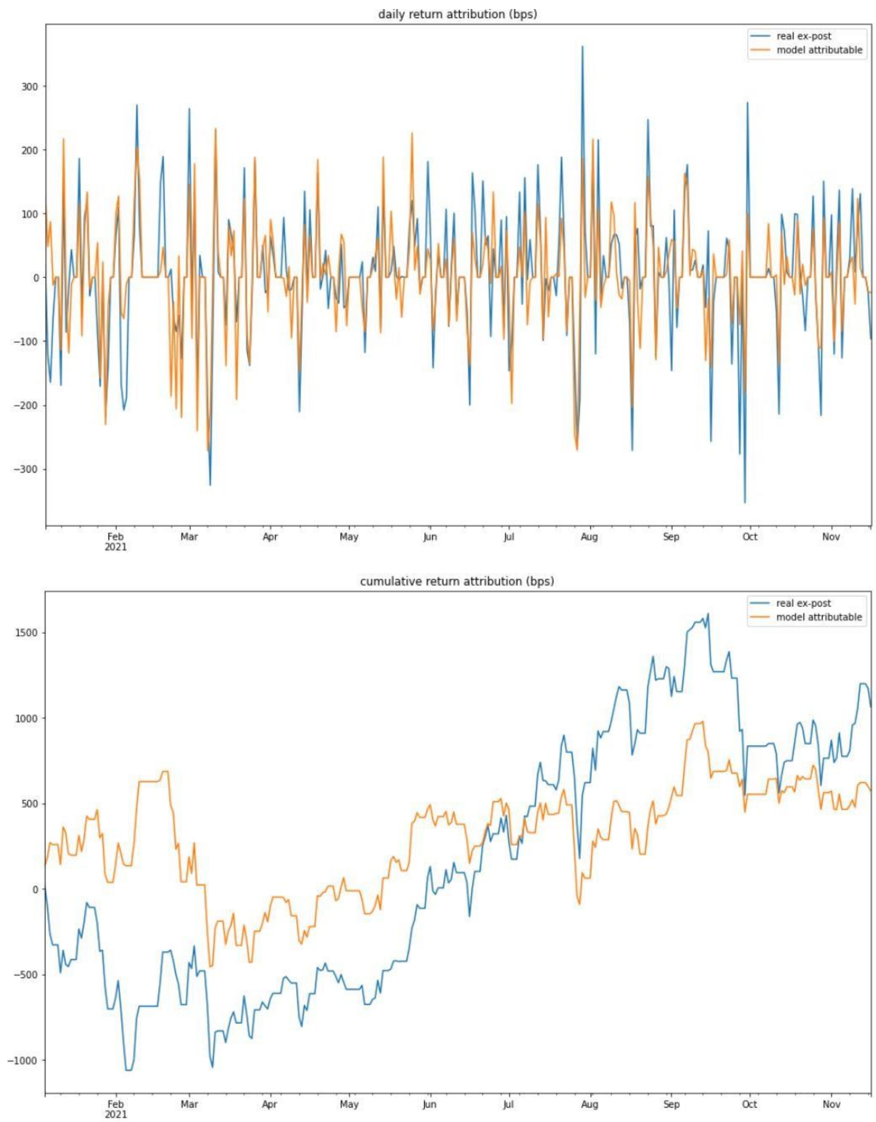
Things are becoming much clearer if looking at the WLS plot above, in bull market, your momentum FP resembles your L/S beta FP. In bear market, these two deviate from each other.



I am grateful for valuable insights gained from countless informal communication with quant equity practitioners from academia and industry, either within the organization (Dr. V Chandrashekaran) or beyond (Dr. W Choi), to name a few.

It is a good exercise to look into the industry return dispersion once we add the industry-return constraint to avoid multi-collinearity. Imaging the market-level industry return expectation is the y=0 line, anything above it means outperformance and vice versa. Hence, 2021 seems like a "value" year for this given country.

As the second test case, with interest in the small-cap market, I calculated the small-cap market portfolio exposure to various risk factors from the WLS with constraint, using market cap as the weight to seek a relatively low turnover profile. Then, I combined it with the FP returns to get the part of returns attributed to our risk model for this LO portfolio. The difference between the attributable part and the realized market cap weighted portfolio returns is the idiosyncratic return, to demonstrate how much potential an active investor can make in this space. As shown below, the residual is negative till July 2021, then picked up to the positive territory in second half of the year.



3. Open Questions

I am grateful for valuable insights gained from countless informal communication with quant equity practitioners from academia and industry, either within the organization (Dr. V Chandrashekaran) or beyond (Dr. W Choi), to name a few.

Q1: How to construct the factor portfolio for returns from OLS/WLS?

Without loss of generality, we only focus on the derivation for WLS with *W (n by n)* as the weight matrix. OLS is just a special case where the *W* is an identity matrix.

The factor returns estimated through cross-sectional regression is essentially the return by only holding certain factor portfolio (FP hereafter). These FPs have unit exposure to the chosen factor, while no exposure to the other factors.

Recall that the algebraic solution of a WLS regression is as follows:

$$\mathrm{f} = (X^t W X)^{-1} X^t W r = S^t r$$

Then, the FP holding *S (n by m)* can be extracted as:

$$\mathrm{S} = W^t X (X^t W X)^{-1}$$

Each column in *S* is the FP holding to a risk factor, where *m* in the number of factor. I will call the *m by m* inverse matrix in the RHS of the equation as *"wrinkle matrix"* according to the west coast convention.

Q2: Could we construct a factor portfolio with zero exposure to other factors using the first approach (long top/short bottom cohort with some tweaks)?

For sure, we could indeed generalize the factor portfolio neutralization from risk factor (e.g. the L/S risk FP we consider here) to any alpha factor.

Say we long the top tertile and short the bottom tertile in the zero-cost portfolio *a*. The top third elements in the *n by 1* vector is positive, the middle third is zero, while the bottom third is negative. We will leverage the above mentioned *"wrinkle matrix"* in this exercise. The key is to disentangle the part of the signal that is coming from the stock's exposure to the different attributes and the forecast raw signal for the FPs corresponding to these attributes.

The forecast raw signal for the FPs, is in the shape of *(m-1 by 1)* since there are *m-1* risk factors left. The notations below will be assumed to automatically adapted to the dimension change:

$$(X^t W X)^{-1} X^t W a$$

Then, we left multiply with the stock exposure matrix to the *m-1* risk factors, resulted in the implied signal adjustment:

$$X (X^t W X)^{-1} X^t W a$$

Finally, we subtract the adjustment from the zero-cost portfolio *a* to get another *n by 1* vector as our neutralized FP.

$$a - X (X^t W X)^{-1} X^t W a$$

4. Quality Control

In the data preparation, we checked the timing of input data. Specifically, on the index universe selection, we confirm taking the matched snapshots of exposure and return respectively after the index reconstitution.

I am grateful for valuable insights gained from countless informal communication with quant equity practitioners from academia and industry, either within the organization (Dr. V Chandrashekaran) or beyond (Dr. W Choi), to name a few.

All the data processing steps are implemented in vectorized python functions. Besides, for those python functions with dimensionality as a default parameter, we explicitly clarify it across the program. It will help a lot if dealing with a 4 or higher dimensional tensor when building a neural network, say LSTM.

In the factor return estimation function, we stick to the for loop structure for better readability and much flexibility in the numerical estimation. A specific quality check is applied to the constraint of our optimization to ensure the industry-level aggregate return is zero within 1 bps tolerance.

In real life, risk model is usually updated on daily/weekly basis using the new coming data point, which will relief much computational time. On the other side, there is some tricky proxying issue we did not meet here. For a security without nearest neighbor data in the universe, sedol proxing may be applied to map it to another one with the same issuer. Finally, it is an open debate if we should explicitly restrict some factor exposure of the market portfolio to be zero in cross-section, which is left for further discretionary decision.

Reference:

[1] Grinold R C, Kahn R N. Active portfolio management[J]. 2000.

[2] Grinold R C. Attribution[J]. The Journal of Portfolio Management, 2006, 32(2): 9-22.

[3] Pástor Ľ, Stambaugh R F, Taylor L A. Dissecting green returns[J]. Journal of Financial Economics, 2022, 146(2): 403-424.

I am grateful for valuable insights gained from countless informal communication with quant equity practitioners from academia and industry, either within the organization (Dr. V Chandrashekaran) or beyond (Dr. W Choi), to name a few.