Asset Level CMAs

Arnav Sheth
iCapital Network

February 29, 2024

Summary

Methodology

Haddad et al, Review of Financial Studies 33(2020): pp.1980-2018

1. Start from a set of pricing factors F_{t+1}.
2. Reduce this set of factors to a few dominant components, Z_{t+1}, using principal components analysis.
3. Produce separate individual forecasts of each of the Z_{t+1}, that is measures of $\mathbb{E}_t[Z_{t+1}]$.
4. To measure the conditional expected factor returns, apply these forecasts to factors using their loadings on the dominant components.
5. To engage in factor timing or estimate the SDF, use these forecasts to construct the portfolio given in Equation (10).

Methodology

Haddad et al, Review of Financial Studies 33(2020): pp.1980-2018

1. Start from a set of pricing factors F_{t+1}.
2. Reduce this set of factors to a few dominant components, Z_{t+1}, using principal components analysis.
3. Produce separate individual forecasts of each of the Z_{t+1}, that is measures of $\mathbb{E}_t[Z_{t+1}]$.
4. To measure the conditional expected factor returns, apply these forecasts to factors using their loadings on the dominant components.
5. ~~To engage in factor timing or estimate the SDF, use these forecasts to construct the portfolio given in Equation (10).~~

Methodology

- Use ~~fifty~~ fifty-five ‘anomaly’ portfolios from Kozak et al. (2020)[1] that effectively capture market heterogeneity.
    - These anomalies are the usual anomalies like Size, Value, ROA, SUE, etc.
- Break them into deciles, create long-short portfolios for each anomaly (Decile 10 minus Decile 1).
    - For each portfolio, they calculate the market-cap-weighted book-to-market ratio (bm) of the underlying stocks.
    - By finding the difference in log book-to-market of Portfolio 10 minus that of Portfolio 1.
- NOTE: This is a placeholder until we sort out data issues relating to ‘CMAs’ (indexes).

[1] Kozak, S., S. Nagel, and S. Santosh. 2020. Shrinking the cross-section. *Journal of Financial Economics* 135:271–92.

Methodology

- Market-adjust and rescale the data.
  1. Calculate regression β for each anomaly.
  2. Market-adjust returns and predictors by subtracting $\beta \times r_{mkt}$ for returns and $\beta \times bm_{mkt}$ for bm ratios.
  3. Rescale to equalize the variance of market-adjiusted returns, and bm ratios for each anomaly.
- So now they have ~~50~~ 55 long-short portfolios.

Methodology

- They conduct a PCA to reduce the ~~50~~ 55 long-short portfolios to five PCs that explain roughly 60% of the variance.

Table 1
Percentage of variance explained by anomaly PCs

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
% var. explained	25.8	12.4	10.3	6.7	4.8	4.0	3.6	2.8	2.2	2.1
Cumulative	25.8	38.3	48.5	55.2	60.0	64.0	67.6	70.4	72.6	74.7

This table reports the percentage of variance explained by each PC of the fifty anomaly strategies.

- Why five?
  1. Campbell and Thompson (2007)[2] show that the monthly R^2 when predicting the market is around 75bp.
  2. A loose upper-bound on the annual Sharpe is 1, or 8.3% monthly.
- If each included PC contributes equally to the R^2, the harmonic mean of their contribution to the total variance of returns must be $> \frac{0.75}{8.3} \approx 9\%$.

[2] Campbell, J. Y., and S. B. Thompson. 2007. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21:1509–31.

Methodology

We extend the Haddad et al. study in the following way:

1. Asset class reduction using PCA/SVD.
2. Regression analysis of principal components on factors (store coefficients).
3. Calculate expected returns of factors.
4. ToDo:
    1. Calculate variance-covariance matrix of factors using regression coefficients, historical data, and residuals.
        - Assume serial correlation.
    2. Use asset-level factor loadings from G3 to get asset-level CMAs.
    3. Myriad data issues.

Literature

Factor Models

- The original one-factor model is the CAPM (Treynor (1961, 1962), Sharpe (1964), Lintner (1965) and Mossin (1966)).
- This was followed of course by Fama-French's three-factor (1993), and five-factor (2014) model.
- Between 1961 and the present, several hundred factors have been used to explain market inefficiency, market risk, and more.
    - This led to the phenomenon commonly termed the 'factor zoo', or the numerous factors used to explain stock market returns.

Literature

Factor Zoo

- What is the 'factor zoo'?
  - First mentioned in Harvey, Liu, and Zhang (2016) following McLean and Pontiff (2015).
  - Main issue is the proliferation of factors – they test 316!
  - They find that many of these factors are highly correlated.
- So dimensionality reduction is highly recommended. This applies to our CMA problem as well.
  - Envestnet, for example, has 55 asset classes.

Literature

PCA

- The use of PCA in asset pricing goes back to a series of papers by Connor and Korajczyk (1986, 1988, 1993).
- More recently, Kozak et al. (2020) have done work on dimensionality reduction of the cross-section of asset returns.
- Several other papers include: Kozak and Nagel (2023), Nadler and Sancetta (2023), and more[3].

[3] If you're interested, Fama and Stern (2016) have an interesting conversation on factor calculation using cross-sectional versus time series data.

Data

- We start with the anomaly dataset used in Haddad, Kozak, Santosh (2020) and Giglio, Kelly, Kozak (2020).
- Dataset consists of decile data along 55 equity anomalies.
    - These consist of characteristics like size, value, momentum, reversal, etc.
- Why this dataset?
    - It is easily accessible, and organized well.
    - It serves as a good placeholder for CMA data.
    - It is the same database used in Haddad et al.

Data

Anomaly Dataset (Subset)

1. **Size** (*size*). Follows Fama and French (1993). size = $\mathrm{ME}_{\mathrm{Jun}}$. The CRSP end of June price times shares outstanding. Rebalanced annually.
2. **Value (annual)** (*value*). Follows Fama and French (1993). value = BE/ME. At the end of June of each year, we use book equity from the previous fiscal year and market equity from December of the previous year. Rebalanced annually.
3. **Gross Profitability** (*prof*). Follows Novy Marx (2013a). prof = GP/AT, where GP is gross profits and AT is total assets. Rebalanced annually.
4. **Value-Profitability** (*valprof*). Follows Novy Marx (2013b). valprof = rank(value) + rank(prof). Sum of ranks in univariate sorts on book-to-market and profitability. Annual book-to-market and profitability values are used for the entire year. Rebalanced monthly.
5. **Piotroski's *F*-score** (*F-score*). Follows Piotroski (2000). F-score = $1_{\mathrm{IB}>0}+1_{\Delta\mathrm{ROA}>0}+1_{\mathrm{CFO}>0}+1_{\mathrm{CFO}>\mathrm{IB}}+1_{\Delta\mathrm{DTA}<0|\mathrm{DLTT}=0|\mathrm{DLTT}_{-12}=0}+1_{\Delta\mathrm{ATL}>0}+1_{\mathrm{EqIss}\leq 0}+1_{\Delta\mathrm{GM}>0}+1_{\Delta\mathrm{ATO}>0}$, where IB is income before extraordinary items, ROA is income before extraordinary items scaled by lagged total assets, CFO is cash flow from operations, DTA is total long-term debt scaled by total assets, DLTT is total long-term debt, ATL is total current assets scaled by total current liabilities, EqIss is the difference between sales of of common stock and purchases of common stock recorded on the cash flow statement, GM equals one minus the ratio of cost of goods sold and total revenues, and ATO equals total revenues, scaled by total assets. Rebalanced annualy.
6. **Debt Issuance** (*debtiss*). Follows Spiess and Affleck-Graves (1999). debtiss = $1_{\mathrm{DLTISS}\leq 0}$. Binary variable equal to one if long-term debt issuance indicated in statement of cash flow. Updated annually.

Start date: 07-01-2004. End date: 2019-12-01. Size: 186 × 54.

Data

Factor Data

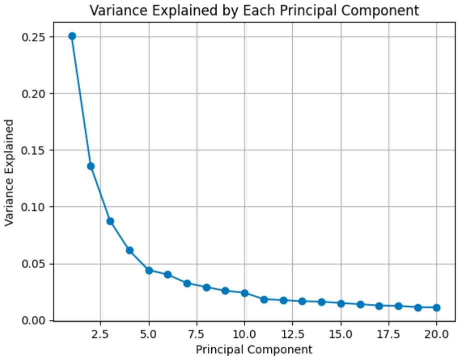
Factor Name	Index
Alt Commodities	Bloomberg Commodity Index
Alt HF Crowding	*Difference Between:* Barclays Form 13 Filing; Russell 3000 Total Return Index
Alt Oil	*Average of:* Middle East Crude Oil; WTI Crude Oil; Brent Crude Oil
Alt Trend	Credit Suisse Managed Futures Index
Emerging Markets	*Difference Between:* MSCI Emerging Markets Index (PR); MSCI ACWI (PR)
Equity Market	MSCI ACWI (TR)
Equity Momentum	*Difference Between:* MSCI ACWI Momentum NR USD; MSCI ACWI IMI
Equity Quality	*Difference Between:* MSCI ACWI Quality NR USD; MSCI ACWI IMI
Equity SmallCap	*Difference Between:* S&P BMI Global Small-cap PR; S&P Global Broad Market PR
Equity Value	*Difference Between:* S&P BMI Global Value PR; S&P Global Broad Market PR
Fixed Credit	*Difference Between:* Bloomberg US Corporate High Yield Index; Bloomberg Barclays US Aggregate Bond Index
Fixed Duration	Bloomberg U.S. Treasury: 7-10 Year Total Return Index Value Unhedged
US Dollar	US Dollar Index

Start date: 07-01-2004. End date: 2019-12-01. Size: 186 × 13.

Results[4]

Caveat: these results should be used for idea generation and for code plcaeholders; not taken literally.

- Let's start with the dimensionality reduction for the anomalies dataset.



Scree plot

[4] Caveat: these results should be used for idea generation and for code plcaeholders; not taken literally.

Results[5]

Principal Component	Explained Variance (%)	Cumulative Variance (%)
1	25.091289	25.091289
2	13.603143	38.694431
3	8.774071	47.468503
4	6.150112	53.618614
5	4.415533	58.034148
6	4.014095	62.048243
7	3.257016	65.305259
8	2.913916	68.219175
9	2.580170	70.799344
10	2.408403	73.207747
11	1.844172	75.051919
12	1.748976	76.800895
13	1.660144	78.461039
14	1.616413	80.077452
15	1.499114	81.576566
16	1.393372	82.969938
17	1.279384	84.249322
18	1.243551	85.492873
19	1.133469	86.626342
20	1.116908	87.743249

- Roughly 80% of the variance is explained by 14 principal components.

[5] Caveat: these results should be used for idea generation and for code plcaeholders; not taken literally.

Results[6]

- Running a OLS regression with HAC-robust standard errors, we get, for Alt Commodities:

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.2537	0.327	-0.776	0.439	-0.899	0.392
PC1	0.0729	0.089	0.820	0.413	-0.102	0.248
PC2	0.0380	0.121	0.315	0.753	-0.200	0.276
PC3	-0.0626	0.150	-0.417	0.677	-0.359	0.234
PC4	-0.2105	0.179	-1.173	0.242	-0.565	0.144
PC5	0.0696	0.212	0.329	0.743	-0.348	0.488
PC6	-0.5006	0.222	-2.254	0.025	-0.939	-0.062
PC7	-0.1688	0.247	-0.685	0.495	-0.656	0.318
PC8	0.4181	0.261	1.604	0.111	-0.097	0.933
PC9	0.0365	0.277	0.132	0.895	-0.510	0.583
PC10	-0.4101	0.287	-1.430	0.155	-0.976	0.156
PC11	-0.9705	0.328	-2.961	0.003	-1.617	-0.324
PC12	-0.1718	0.337	-0.510	0.610	-0.836	0.493
PC13	-0.2090	0.345	-0.605	0.546	-0.891	0.473
PC14	0.8251	0.350	2.357	0.020	0.134	1.516

Omnibus:	3.574	Durbin-Watson:	2.014
Prob(Omnibus):	0.167	Jarque-Bera (JB):	4.068
Skew:	-0.082	Prob(JB):	0.131
Kurtosis:	3.706	Cond. No.	3.94

[6] Caveat: these results should be used for idea generation and for code plcaeholders; not taken literally.

Results[7]

- Perhaps run a stepwise regression instead?

```
Stepwise regression for dependent variable Alt Commodities:
Selected features: ['PC11', 'PC14', 'PC6']


Stepwise regression for dependent variable Alt HF Crowding:
Selected features: ['PC10', 'PC13', 'PC11', 'PC3', 'PC7']


Stepwise regression for dependent variable Alt Oil:
Selected features: ['PC6', 'PC11', 'PC1']


Stepwise regression for dependent variable Alt Trend:
Selected features: []
```

[7] Caveat: these results should be used for idea generation and for code plcaeholders; not taken literally.

Issues

- How do we use the CMA expected returns to 'plug' into our regression results?
    - Reconstructing the original matrix using reverse-SVD results in a high MSE[8].
        - Rabbit-hole: We can use Frobenius norm-minimization techniques for reconstruction.
    - Take the median of the respective PCs as substitutes for the historical expected returns? (Simplest solution, IMHO).
    - Use the means of the original dataset and run another PCA with the same number of components?
- Do we even want to follow this approach of using external CMAs to generate asset-level CMAs?

[8] Thanks, Aniket, for your help here.

Issues

- How do we generate variance-covariance matrices using the PCA plus regression approach?
    - We must account not only for the coefficient uncertainty but also the contribution to variance of each individual PC.