Smart Beta Trading Strategy Based on Fundamental Factors

Nan Xi 20666874; Zhao Yanan 20671764; Shi Xinrun 20660076; Ji chunhua 20567626

Abstract

In our report, fundamental factors based on Fama-French model are selected to construct the strategy. Stocks in US market is used as the universe, and SPY in quantopian is used as the benchmark. In the first part of our report, we run single-factor back-testing and OLS regression to choose factors related to stock return. Then we conduct hedging based on the OLS regression model we choose in step one. In the second part, we combine the selected factors in our to conduct long-short strategy based on z-score of different stocks. Finally, we compare the cumulative return of our portfolio with benchmark to certificate the effectiveness of our strategy.

1. Background

In recent years people have found that actively seeking alpha make not be better off than just passively tracking index. PMs can barely provide sustainable alpha strategy. Furthermore, since a proactive strategy may also bring in too much tracking error from specific sectors or companies, it is so not welcomed by asset allocation purpose. This is why ETFs are getting more popular. But just passive index tracking strategy may not work so well when the market is weak. There is no diversification on the systematic risk and no specific risk exposure to get extra return. A so-called smart beta strategy may still be able to keep the balance between return and risk.

Since the CAPM model was first developed in 1950s, there are tons of comprehensive and advanced models developed afterwards, trying to decompose stock returns. Much research has been done on determining significant factors, and what makes things even more difficult is that the discovery of a significant factor often leads to its advantage being arbitraged away. As time goes by, most of these well researched factors lost their edges as market efficiency improves. This is one of the reasons why fundamental factor models, and linear factor models in general, are so prevalent in modern finance.

Fundamental data refers to the metrics and ratios measuring the financial characteristics of companies derived from the public filings made by these companies, such as their income statements and balance sheets. Examples of factors drawn from these documents include market cap, net income growth, and cash flow.

This fundamental data can be used in many ways, one of which is to build a linear factor model. Given a set of k fundamental factors, we can represent the returns of an asset as follows:

$$R_t = \alpha_t + \beta_{t,F_1} F_1 + \beta_{t,F_2} F_2 + \ldots + \beta_{t,F_k} F_k + \epsilon_t$$

where each F represents a fundamental factor return stream. These return streams are from portfolios whose value is derived from it's respective factor.

Fundamental factor models try to determine characteristics that affect an asset's risk and return. The most difficult part of this is determining which factors to use. Once we have found significant factors, we need to calculate the exposure an asset's return stream has to each factor.

2. Simple-factor Trading Strategy

2.1 factor introduction

The first approach consists of using the fundamental data as a ranking scheme and creating a long-short equity portfolio based on each factor. We then use the return streams associated with each portfolio as our model factors.

Historically, certain groups of stocks were seen as outperforming the market, namely those with small market caps, high book-to-price ratios, and those that had previously done well (i.e., they had momentum). Empirically, Fama & French found that the returns of these particular types of stocks tended to be better than what was predicted by the security market line of the CAPM. In order to capture these phenomena, we will use those factors to create a ranking scheme that will be used in the creation of long short equity portfolios.

Our research base on classic five-factor Fama-French model. The factors will be SMB, measuring the excess return of small market cap companies minus big; HML, measuring the excess return of companies with high book-to-price ratios versus low ; EXMRKT which is a measure of the market risk; RMW, measuring the excess return of enterprise with robust profitability against weak; CMA, measuring the excess return of enterprise with aggressive reinvestment strategy against conservative.

And in order to improve the model, we add another fundamental factor: MOM, measuring the excess returns of last month's winners versus last month's losers. Our universe is US500 collected in the database in quantopian. We collected the following data in 2018, because we do not have compete database in quantopian in 2019.

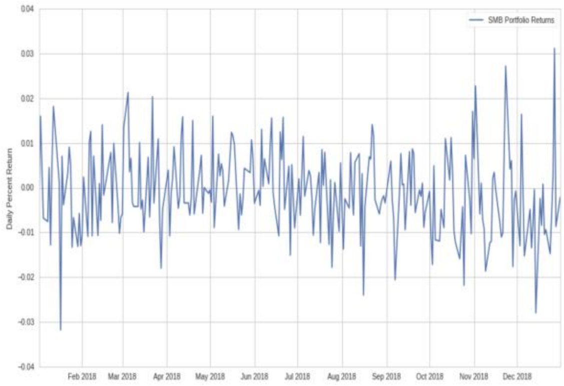
factor	proxy
EXMRKT	SPY-BIL
SMB	Market Value - using Period End Price
HML	Book Value Per Share/ close price
RMW	Return on Average Total Equity
CMA	reinvestment rate
MOM	Return of monthly close price

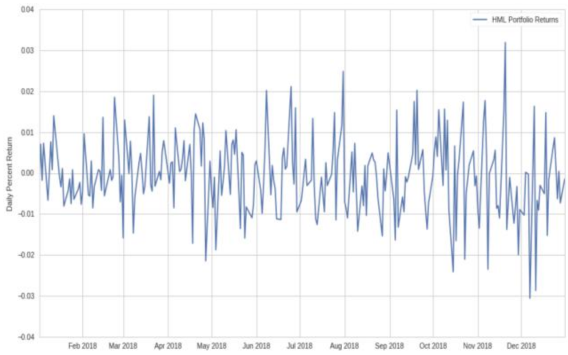
(from quantopian FactSet Fundamental).

2.2 Methodology and findings

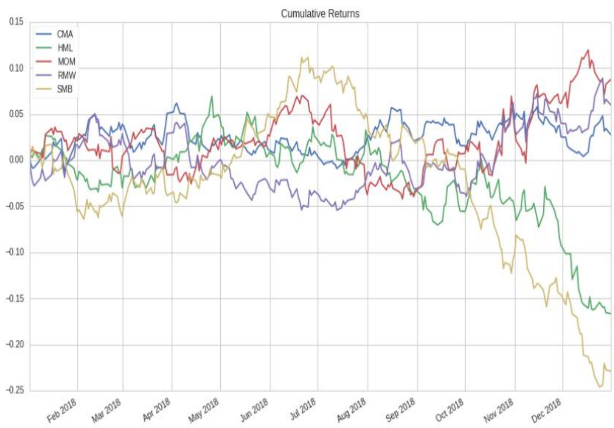
First, we compute a daily rank of these six factors, this is used in the next step, which is computing portfolio membership. Then we grab the top and bottom 50 for each factor and define our universe, screening out anything that isn't in the top or bottom. After initializing the pipe, we can go through the data and build the factor portfolios we want. After constructing our portfolios, we can check our performance if we were to hold each one.

Result analysis:





The cumulative return from these factors is as follows:



As we can see from the chart above, cumulative return of SMB and HML did not perform well in US market, losing more than 15% of total capital in 2018. On the other hand, momentum, RMW and CMA gave us positive return in 2018. As we know, the classic three-factor model was given by Fama & French in 1993. Size premium and value premium may not exist in today's market any more. Moreover the model is largely based on CAPM, which has many unrealistic hypotheses. Also, the realization of the model may because of data mining bias to some extend. In this way, in the following part of the research, we will not take these two factors into account. The back test results of momentum, RMW and CMA are given:

Momentum

	Backtest
Annual return	8.783%
Cumulative returns	8.746%
Annual volatility	15.641%
Sharpe ratio	0.62
Calmar ratio	0.84
Stability	0.10
Max drawdown	-10.464%
Omega ratio	1.11
Sortino ratio	0.93
Skew	0.45
Kurtosis	3.29
Tail ratio	0.97
Daily value at risk	-1.932%
Alpha	0.10
Beta	0.03

RMW

	Backtest
Annual return	6.031%
Cumulative returns	6.007%
Annual volatility	13.649%
Sharpe ratio	0.50
Calmar ratio	0.61
Stability	0.09
Max drawdown	-9.869%
Omega ratio	1.08
Sortino ratio	0.73
Skew	0.02
Kurtosis	0.47
Tail ratio	1.12
Daily value at risk	-1.693%
Alpha	0.07
Beta	-0.02

CMA

	Backtest
Annual return	2.795%
Cumulative returns	2.783%
Annual volatility	10.644%
Sharpe ratio	0.31
Calmar ratio	0.42
Stability	0.09
Max drawdown	-6.679%
Omega ratio	1.05
Sortino ratio	0.43
Skew	-0.39
Kurtosis	0.75
Tail ratio	0.98
Daily value at risk	-1.328%
Alpha	0.03
Beta	-0.00

Among these three factors, momentum has the largest cumulative return, Sharpe ratio, Calmar ratio etc. But it also has the maximum drawdown, VaR. On the one hand, momentum gives the best performance; on the other hand, it has the maximum volatility. Next step, we'll calculate the risk premium on each of these factors using the OLS regression.

OLS Regression Results

Dep. Variable:	Returns	R-squared:	0.069
Model:	OLS	Adj. R-squared:	0.059
Method:	Least Squares	F-statistic:	6.987
Date:	Sun, 31 May 2020	Prob (F-statistic):	3.50e-07
Time:	05:43:06	Log-Likelihood:	2625.3
No. Observations:	572	AIC:	-5237.
Df Residuals:	565	BIC:	-5206.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
const	-8.07e-05	0.000	-0.526	0.599	-0.000 0.000
EXMRKT	-0.0003	0.000	-1.179	0.239	-0.001 0.000
SMB	-0.0005	0.000	-3.319	0.001	-0.001 -0.000
HML	-3.143e-05	0.000	-0.202	0.840	-0.000 0.000
MOM	0.0008	0.000	4.776	0.000	0.000 0.001
RMW	0.0005	0.000	3.454	0.001	0.000 0.001
CMA	0.0004	0.000	3.537	0.000	0.000 0.001

Omnibus:	105.450	Durbin-Watson:	2.058
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1543.952
Skew:	0.276	Prob(JB):	0.00
Kurtosis:	11.030	Cond. No.	8.84

The OLS regression results show that market factor and HML are not significant. MOM, RMW and CMA have significant positive relationship with return. SMB however, is negative related to the return, which is not rational. P-value of the model is obviously smaller than 5%, meaning the model is significant. $R^2 = 0.069$, meaning 6.9 percent of return can be explained by theses five factors.

Factor modeling can be used to predict future returns based on current fundamental factors. As well, it could be used to determine when an asset may be mispriced in order to arbitrage the difference. Modeling future returns is accomplished by offsetting the returns in the regression, so that rather than predict for current returns, we are predicting for future returns. Once we have a predictive model, the most canonical way to create a strategy is to attempt a long-short equity approach.

2.3 Portfolio Construction for Hedging

Other than use this approach as an asset pricing model we can also use it to hedge our portfolios. Hedging can be achieved through a linear regression of portfolio returns on the returns from the long-short factor portfolios. Once we've determined that we are exposed to a factor, we may want to avoid depending on the performance of that factor by taking out a hedge. The essential idea is to take the exposure our return stream has to a factor, and short the proportional value. So, if total portfolio value was V, and the exposure our calculated to a certain factor return stream was β, we would short βV amount of that return stream.

We'll take a random sample of 50 assets from Q500US in order to build a random portfolio and calculate the return of our portfolio with no hedging. Next, we calculate the exposure of our portfolio to each of the Fama-French factors.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                Returns   R-squared:                       0.12
Model:                            OLS   Adj. R-squared:                  0.12
Method:                 Least Squares   F-statistic:                     234.
Date:                Sun, 31 May 2020   Prob (F-statistic):          2.79e-18
Time:                        06:36:32   Log-Likelihood:                 16502
No. Observations:                6658   AIC:                        -3.299e+0
Df Residuals:                    6653   BIC:                        -3.296e+0
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.
------------------------------------------------------------------------------
EXMRKT         0.0082      0.023      0.352      0.725        -0.038     0.05
MOM           -0.0732      0.026     -2.812      0.005        -0.124    -0.02
RMW           -1.6119      0.063    -25.468      0.000        -1.736    -1.48
CMA            1.2815      0.081     15.876      0.000         1.123     1.44
const      -5.976e-05      0.000     -0.240      0.810        -0.001     0.00
==============================================================================
Omnibus:                     1574.599   Durbin-Watson:                   1.63
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            87917.00
Skew:                           0.135   Prob(JB):                         0.0
Kurtosis:                      20.800   Cond. No.                         401
==============================================================================
```

Check for normality, homoskedasticity, and autocorrelation:

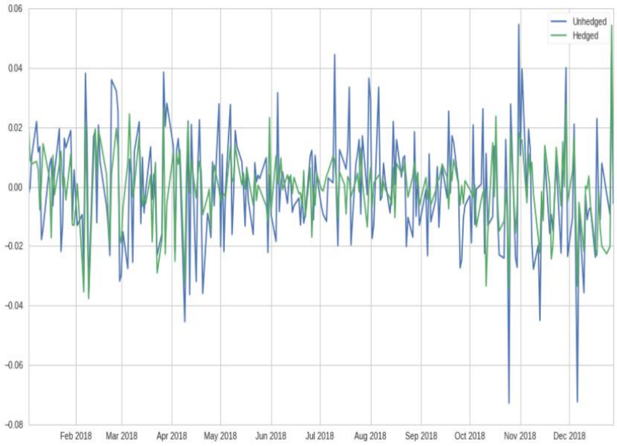
```
Jarque-Bera p-value:  0.0
Breush Pagan p-value:  0.382984555295
Durbin Watson statistic:  1.63525544134
```

For normality, we'll run a Jarque-Bera test, which tests whether our data's skew/kurtosis matches that of a normal distribution. p-value<5%, meaning we would reject the null hypothesis that the data is normally distributed. This means there is strong evidence that our data follows some other distribution.

To test for heteroskedasticity, we'll run a Breush-Pagan test, which tests whether the variance of the errors in a linear regression is related to the values of the independent variables. In this case, our null hypothesis is that the data is homoskedastic. In this case, our data is not heteroskedastic.

Autocorrelation is tested for using the Durbin-Watson statistic, which looks at the lagged relationship between the errors in a regression. We cannot reject the null hypothesis. There is no autocorelation in our data.

Let's look at a graph of our two portfolio return streams:



As we can see, the hedged portfolio has less volatility than the unhedged one, meaning our hedged method can reduce exposed risk in a good way.

3. Multi-factor Strategy

Another approach is to normalize factor values for each day and see how predictive of that day's returns they were. We do this by computing a normalized factor value $b_{a,j}$ for each asset a in the following way.

$$b_{a,j} = \frac{F_{a,j} - \mu_{F_j}}{\sigma_{F_j}}$$

F is the value of factor j for asset a during this time, μ is the mean factor value across all assets, and σ is the standard deviation of factor values over all assets. Notice that we are just doing the standardization to make asset specific factor values comparable across different factors. Winzorization takes the top n% of a dataset and sets it all equal to the least extreme value in the top n%.We use this method to replace extreme data values with less extreme values.

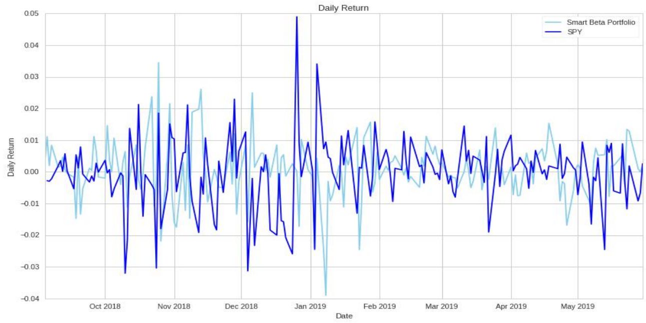
The universe of our final strategy is US1500 collected in quantopian. Benchmark we chosen is SPY. The factors chosen are RMW, CMA and 10-period momentum. Firstly, we standardize each factors of a stock, and then put weights on them. Then, we add these three factors together, getting the total z-score of a stock. Finally, we rank the z-scores of different stocks from high to low, buying 200 stocks with highest z-score, selling 200 stocks with lowest z-score from 2018.9.1 to 2019.5.30. The weight we put on different factors are based on their annual return of simple factor back-testing which we discussed in part one:

factor	weight
RMW	0.34
CMA	0.16
momentum	0.5

We can view our trading strategy of 2018.9.4 as an example:

		Returns	buy	sell	total_z	total_z_rank
2018-09-04 00:00:00+00:00	Equity(110 [RAMP])	0.003956	True	False	0.254912	1328.0
	Equity(168 [AET])	0.003308	True	False	0.236539	1306.0
	Equity(197 [AGCO])	-0.004338	False	True	-0.162957	177.0
	Equity(216 [HES])	-0.012029	False	True	-0.215176	123.0
	Equity(300 [ALK])	0.004614	True	False	0.219956	1287.0

The daily return of our portfolio vs benchmark is:



And the cumulative return is:

As is shown in the chart, our smart Beta portfolio has much higher cumulative return than benchmark, giving us highest return about 14% at the beginning of 2019.

4. Conclusions

(a) Based on the five-factor Fama-French model, we add a fundamental factor: momentum, since it perform well in the recent market. From the analytic results, momentum performs best with the highest volatility, followed by RMW and CMA, while SMB and HML do not give us positive return. OLS regression shows that market factor and HML are not significant , SMB is irrationally negative related to the return. We had better exclude HML and SMB in our model.

(b) Compared with the no-hedged portfolio, hedged one give us positive return with lower volatility.

(c) Finally, we use RMW, CMA and 10-period momentum to multi-factor back-testing. By ranking the stocks with regard to z-score, we perform long-short strategy, and getting excess return against benchmark.

(d) The positive cumulative return we get may be a result of data mining, so we should still be careful when doing real trade.

5. Links

Code: https://github.com/ynnnnnnn/MAFS6010U-final-project.git
Presentation Video: https://youtu.be/VHFIbZrHVno

6. Contributions

ZHAO,Yanan: Discussing strategy, writing code, writing report, presentation
SHI, Xinrun: Discussing strategy, writing code, writing report, presentation
NAN, Xi: Discussing strategy,writing code, writing report, presentation
JI, Chunhua: Discussing strategy, writing report, presentation