Improved Anomalies Strategy

---- Midterm Presentation

Group 4:
Yu Gu
Ruoyu Han
Yuzhu Zhang

Introduction

Background:

- What is market anomaly?

-- when a security or group of securities performs contrary to the notion of efficient market, where security prices are said to reflect all available information at any point in time.

- E.g.

-- small firms / low volatility / high book-to-price stocks tend to outperform

-- January effect

Our questions:

- Are these anomalies still exist, or when will they appear?
- How can we use them?

Problem Definition

Our Goal:

- Detect effective anomalies factors in recent years
- Predict some important factors
- Combine different anomaly signals to construct a portfolio

Current Progress:

- Familiar with the data source (Quantopian)
- Compute 7 classical anomaly factors
- Analyze the performance of these factors

Data Universe

Data Selection

Criterion:

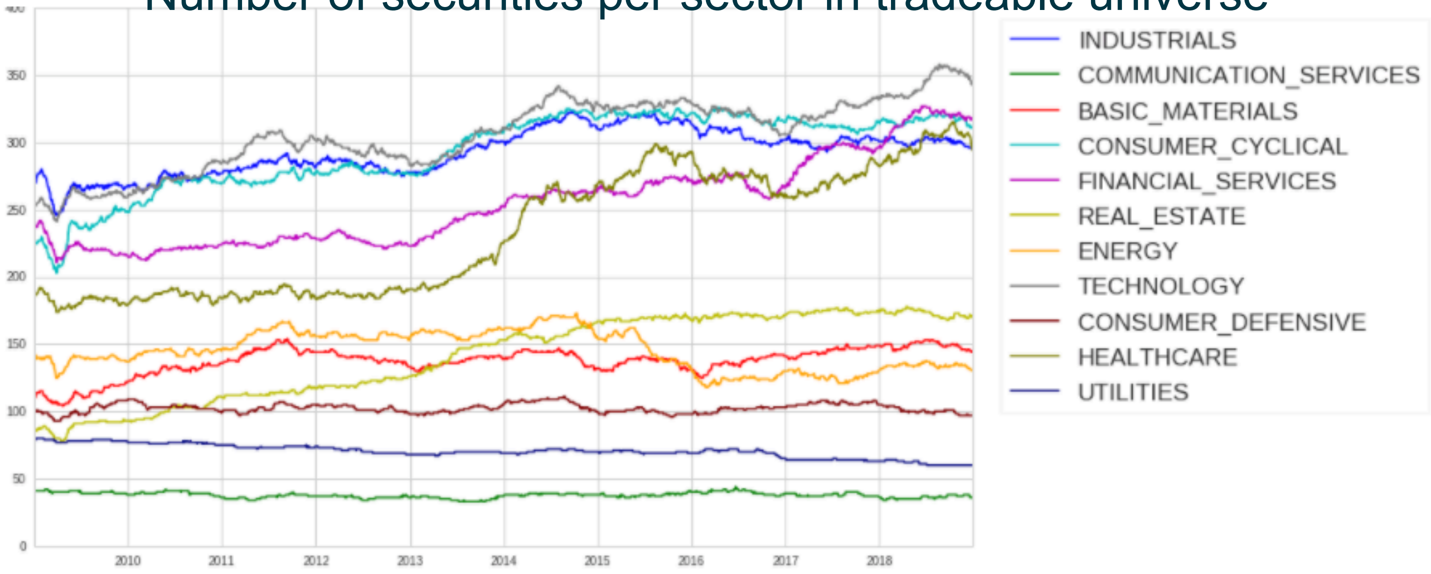
- As much data as possible
- Reliable resource and no survivor bias
- Long time period(10 years period from 2009-1-1 to 2018-12-31)

Stock Universe:

- QTradableStocksUS from Quantopian
- It provides a set of liquid, easy-to-trade stocks while excluding assets that have more difficult risk profiles like ADRs and ETFs
- No explicit size limit, and generally has between 1600-2100 members. We select the top and bottom 500 stocks for each factor
- Remove illiquid or otherwise untradeable stocks
- Updated daily

Number of securities

Number of securities per sector in tradeable universe



Factor Selection

References in literature :

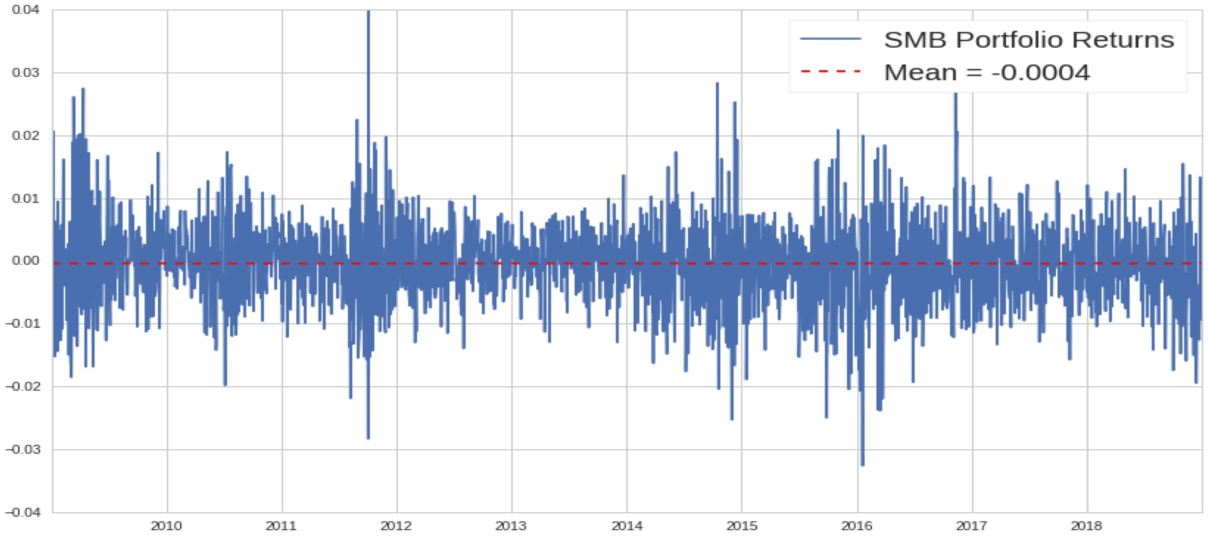
1. Jean-Philippe Bouchaud Pierre Blanc, R´emy Chicheportiche. The fine structure of volatility feedback ii: overnight and intra-day effects. 2014
2. A. Beveratos G. Simon L. Laloux M. Potters J.-P. Bouchaud S. Ciliberti, Y. Lemperiere. Deconstructing the **low-vol** anomaly. 2015
3. Guillaume Simon Yves Lemperiere Jean-Philippe Bouchaud Stefano Ciliberti, Emmanuel Serie. The “**size** premium” in equity markets: Where is the risk?2017
4. Augustin Landier Guillaume Simon Jean-Philippe Bouchaud, Stefano Ciliberti and David Thesmar. The excess returns of “**quality**” stocks: A behavioral anomaly, 2016

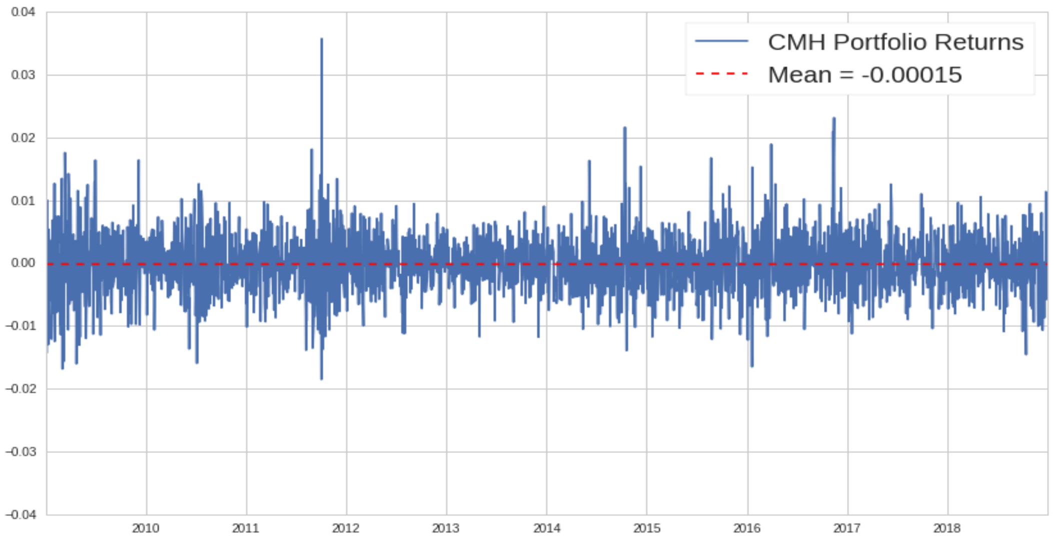
Based on the above four papers, we decided to choose the following factors as a starting point,

Aspects of factors:

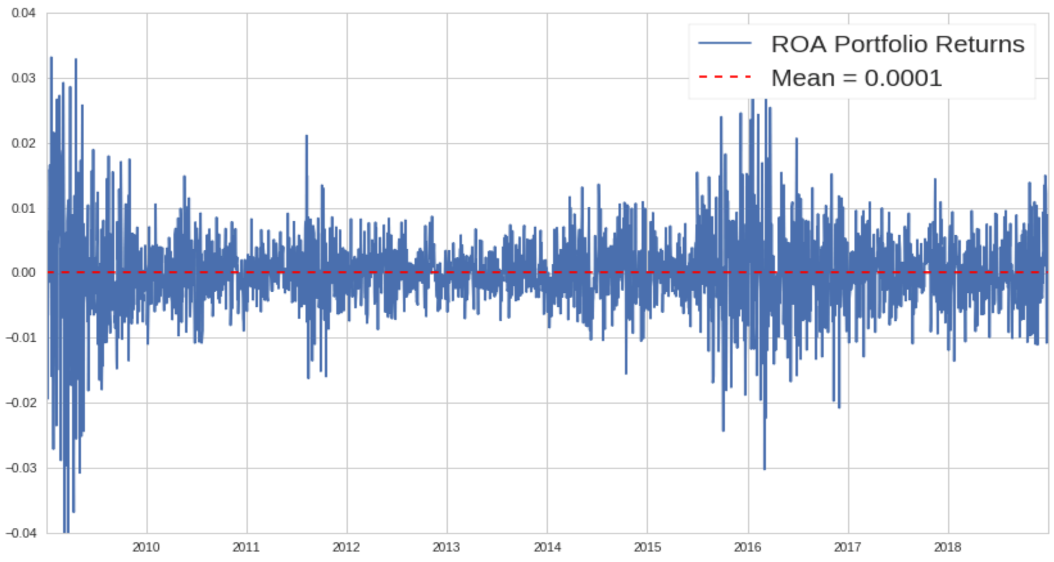
- Size factor: **SMB**(small cap minus big cap), **CMH**(cold minus hot, average daily volume)
- Quality factor: **ROA**(high return-over-assets minus low return-over-assets) , **OCF**(high net operating cash flow minus low net operating cash flow)
- Volatility factor: **LowVol**(low volatility minus high volatility)
- Momentum factor: **UMD**(up minus down momentum)
- Value factor: **HML** (high book-to-price minus low book-to-price)

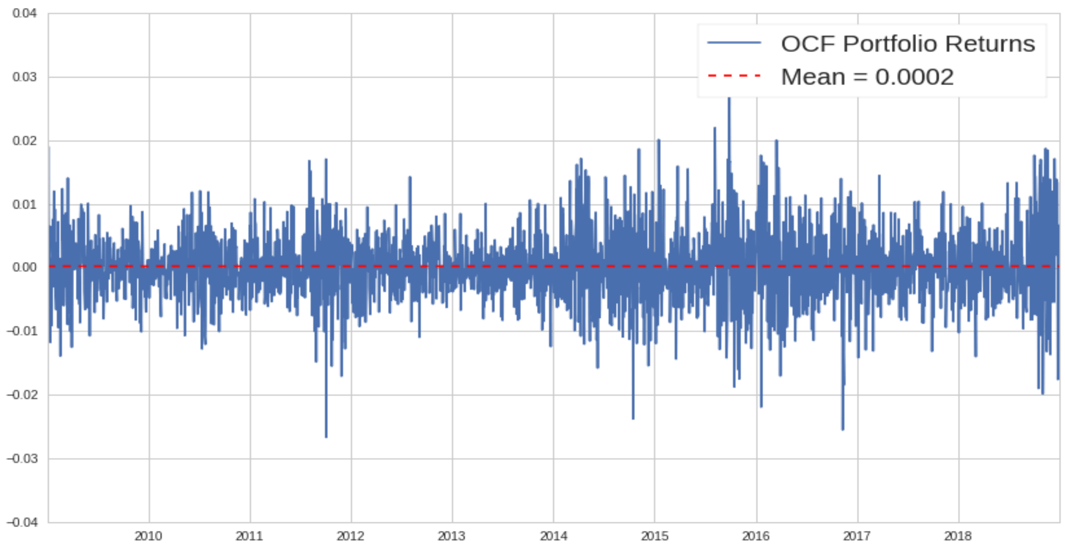
Daily performance of factor portfolios(size factors SMB and CMH):



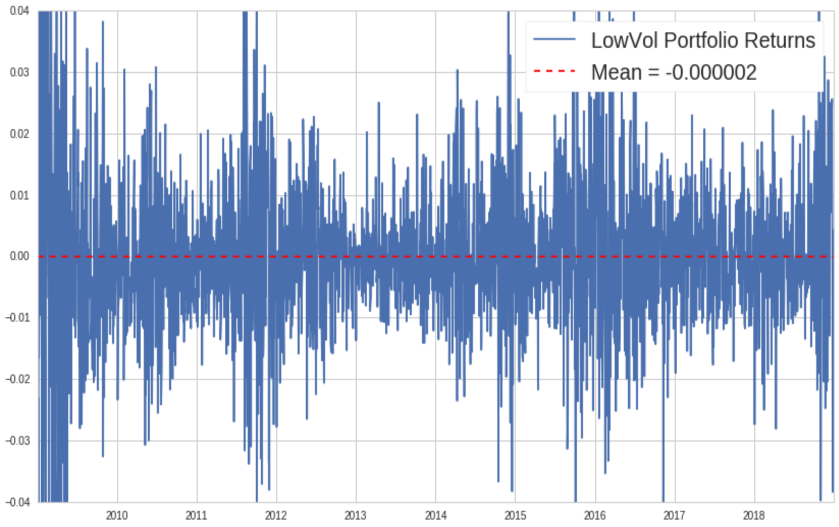


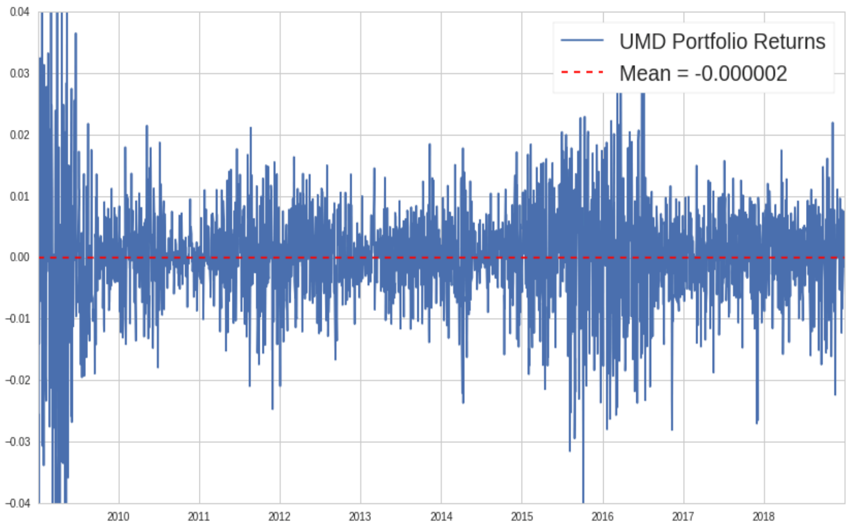
Daily performance of factor portfolios(Quality factor ROA and OCF)

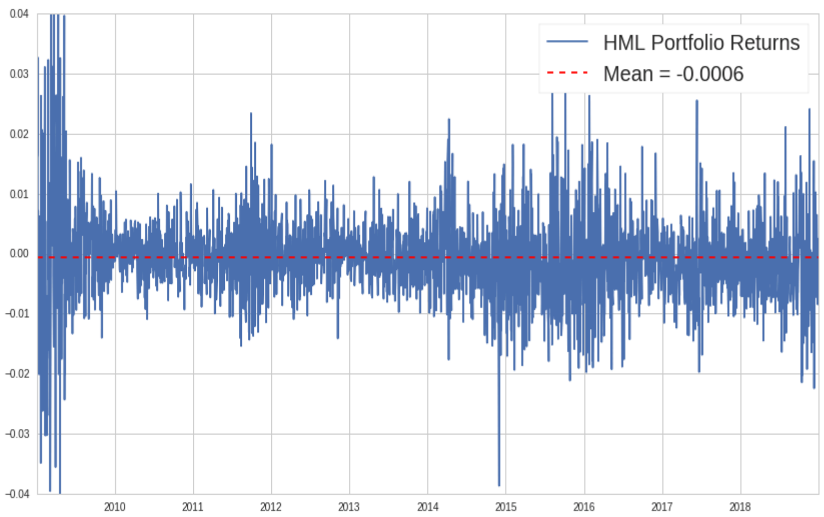




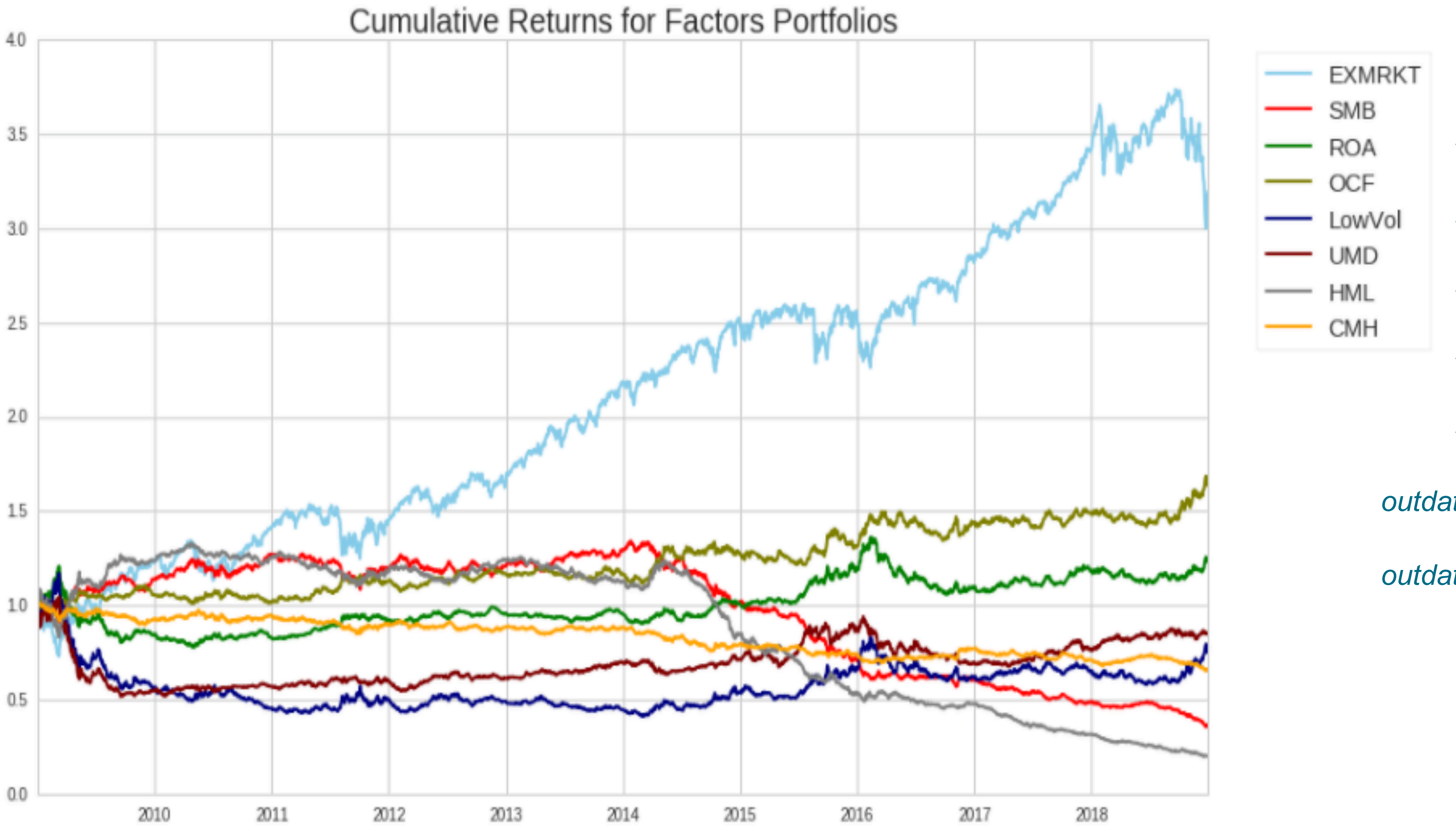
Daily performance of factor portfolios(Volatility factor LowVol, Momentum factor UMD, Value factor: HML)







Cumulative Returns



	EXMRKT	3.199	
*	OCF	1.635	Quality factor
*	ROA	1.234	
*	UMD	0.853	Momentum factor
*	LowVol	0.751	Volatility factor
*	CMH	0.661	Size factor
outdated	SMB	0.363	
outdated	HML	0.198	Value factor

* Factors of interest in further steps

Model Information

OLS Regression results: step 1. estimating the betas for each stock, step 2. estimating the risk premium

Most factors are statistically significant

Risk premium exists

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
const	0.0013	9.56e-05	13.091	0.000	0.001 0.001
EXMRKT	0.0006	0.000	3.958	0.000	0.000 0.001
SMB	-0.0008	6.93e-05	-10.972	0.000	-0.001 -0.001
CMH	-8.229e-05	6.28e-05	-1.311	0.190	-0.000 4.07e-05
ROA	0.0004	4.71e-05	9.317	0.000	0.000 0.001
OCF	0.0007	5.11e-05	14.348	0.000	0.001 0.001
LowVol	0.0015	8.55e-05	17.222	0.000	0.001 0.002
UMD	-0.0004	7.18e-05	-5.332	0.000	-0.001 -0.000
HML	-6.788e-05	6.09e-05	-1.114	0.265	-0.000 5.16e-05

Omnibus:	3294.341	Durbin-Watson:	1.866
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1084575.416
Skew:	-3.273	Prob(JB):	0.00
Kurtosis:	85.824	Cond. No.	13.3

```
Jarque-Bera p-value:  0.0
Breush Pagan p-value:  0.0
Durbin Watson statistic:  1.83567767136
```

Normality_(Jarque-Bera Test)_

p-value of Jarque-Bera test is 0.0, so we would like to reject the null hypothesis that the data is normally distributed. There is strong evidence that our **data follows other distribution.**

Heteroskedasticity_(Breush Pagan Test)_

This tests whether the variance of the errors in a linear regression is related to the values of the independent variables. p-value of Breush Pagan test is also 0.0, suggesting that **the data is heteroskedastic**.

Autocorrelation(_Durbin Watson Test)_

Durbinn Watson statistic test result is 1.84, close to 2. So we cannot reject the null hypothesis of no autocorrelation. Close to 0 will means positive autocorrelation.

Cross-sectional factor analysis

- Another approach. Factor returns are unknows, estimate F_j

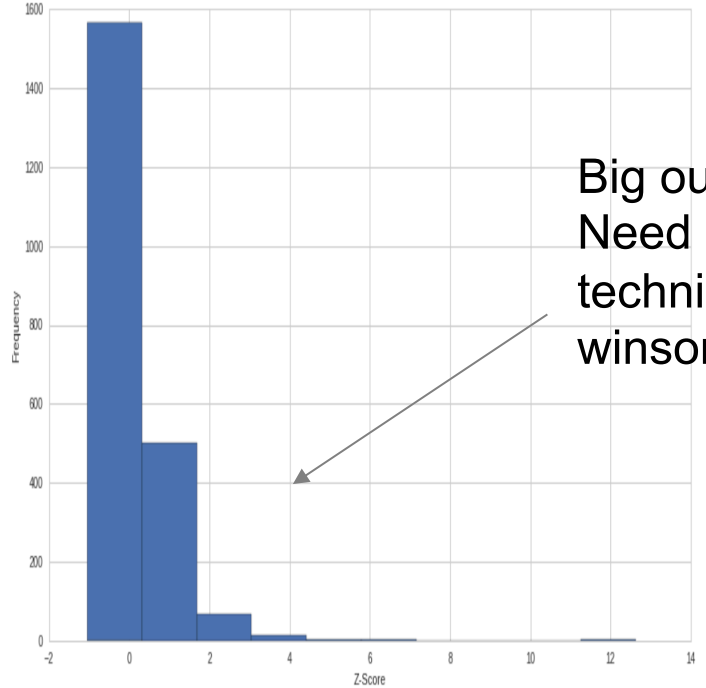
$$R_{a,t} = a_t + b_{a,F1}F_1 + \cdots + b_{a,Fk}F_K$$

- Compute normalized factor value $b_{a,j}$ for each asset a.

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

$\mathrm{F}_{a,j}$ *is the value of factor j for asset a*
μ_{Fj} *is the mean of factor value across all assets*
σ_{Fj} *is the sd of factor value across all assets*

- Another way to think. We are determining how predictive of returns the factor was on that day, and therefore how much return we could have squeezed out of that factor.



Big outliers
Need some cleaning techniques, such as winsorization

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
BTP_z	-0.0039	0.000	-9.979	0.000	-0.005 -0.003
ROA_z	-0.0022	0.000	-5.059	0.000	-0.003 -0.001
MC_z	0.0012	0.001	1.482	0.138	-0.000 0.003
Lag_Ret_z	0.0009	0.000	2.182	0.029	8.77e-05 0.002
CF_z	0.0006	0.000	1.298	0.194	-0.000 0.001
vola_z	-0.0015	0.000	-3.307	0.001	-0.002 -0.001
vol_z	-0.0003	0.001	-0.364	0.716	-0.002 0.001
Constant	0.0053	0.000	13.740	0.000	0.005 0.006

Cross-sectional factor analysis

- Then we loop through days in 2018, and get an estimated factor return

	BTP_z	ROA_z	MC_z	Lag_Ret_z	CF_z	vola_z	vol_z	Constant
2018-01-02	0.000264	0.000983	0.000920	-0.001266	0.000076	0.001797	-0.000693	-0.006452
2018-01-03	0.000688	-0.001850	-0.002477	-0.001841	0.000448	-0.009663	0.003891	0.011459
2018-01-04	-0.001487	-0.001298	-0.001461	0.001413	-0.000057	-0.002098	0.003675	0.002989
2018-01-05	0.001155	0.001829	0.000713	-0.000190	0.000341	0.000607	-0.000409	0.002253
2018-01-08	-0.002188	0.001798	-0.000734	0.001105	-0.000481	0.000435	0.001967	0.002852
2018-01-09	0.000242	0.003976	-0.001865	0.000842	-0.001381	0.002665	0.002354	0.001227
2018-01-10	-0.001329	-0.001662	0.002137	0.002328	0.000843	-0.000362	-0.001304	-0.001771
2018-01-11	0.001036	-0.003263	-0.000087	-0.000640	0.000435	-0.004005	0.001062	-0.000696
2018-01-12	0.001085	0.001702	-0.000998	-0.000704	0.000012	-0.004775	-0.000395	0.014060
2018-01-16	0.000249	0.001596	0.001685	-0.000015	-0.000574	-0.001571	-0.000385	0.004171
2018-01-17	-0.000826	0.002594	0.002422	-0.001848	-0.000096	0.004794	-0.001706	-0.009902

```
n [62]: crosdf.mean()
Out[62]: BTP_z        -5.554314e-04
         ROA_z        -1.642528e-07
         MC_z          1.532806e-04
         Lag_Ret_z     8.047166e-05
         CF_z          4.051046e-06
         vola_z        1.918477e-04
         vol_z        -1.398206e-04
         Constant     -4.451162e-04
         dtype: float64
```

Next Steps

1. Find more predictive anomaly factors in our model
   -- combine factors for the same anomaly
   -- especially the sentiment data of StockTwits

2. Try to apply smarter prediction techniques like machine learning
   -- predict some important anomaly factors and use them to construct portfolios

3. Consider the impact of different sectors
   -- Check the effectiveness of anomaly signals in certain sectors

Thanks!
Q&A

---- Midterm Presentation

Group 4:
Yu Gu
Ruoyu Han
Yuzhu Zhang