Improved Anomalies Strategy

---- Final Presentation

Group 4:
Yu Gu
Ruoyu Han
Yuzhu Zhang

Quick Recap

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?

Project goals:

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

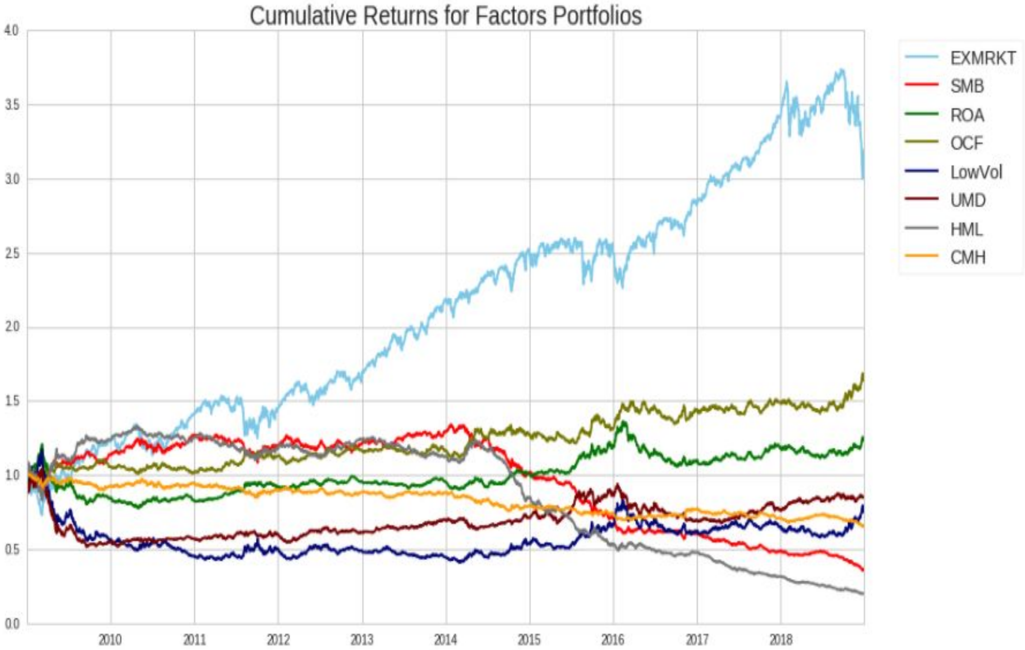
Data universe:

- QTradableStocksUS from Quantopian: reliable resource and no survivor bias
- It provides a set of liquid, easy-to-trade stocks while excluding assets that have more difficult risk profiles like ADRs and ETFs
- Long time period(10 years period from 2009-1-1 to 2018-12-31)

Classical anomalies analysis for midterm

Aspects of factors:

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



Preliminary findings:

- Applied daily factors, the results are not good,
- Factors are time dependent and may only be effective during specific short time period
- Traditional factors like CMH and HML in Famma-French models are outdated

Plan of work after midterm:

- Find more predictive anomaly factors in our model
  -- combine factors for the same anomaly
  -- especially the sentiment data of StockTwits
- Try to apply smarter prediction techniques like machine learning
  -- predict some important anomaly factors and use them to construct portfolios

Revisiting Factors

Aspects of factors:

- Market(size, volatility)
- Profitability
- Growth
- Momentum
- Value
- Liquidity

Market factors	Momentum factors	Value factors	Growth factors	Profit factors	Liquidity factors
size(Market capitalization)	mom6(Momentum in 6 months)	bp(Book to price ratio)	ag(Total assets growth)	roa(Return over assets)	cr(Current ratio)
Beta, betasq(Market beta and its square)	mom12(in 12 months)	ep(Earnings to price ratio)	epsg(EPS growth)	roe(Return over equity)	qr(Quick ratio)
vol(Total volatility)	mom36(in 36 months)	cfp(Cash flow to price ratio)	dpsg(DPS growth)	ato(Asset Turnover)	cf_sale(Cash flow to sales)
skew(Total skewness)	momchg(Momentum change)	sp(Sales to price ratio)	bpsg(BPS growth)	fcf(FCF yield)	
Turn,std_turn(Turnover and its volatility	lagretn(Short term reversal)	peg(PEG ratio)	nig(Net income growth)		
volumed,std_volumed(Dollar volume and its volatility)		ev_ebitda(EV to EBITDA)	oig(Operating income growth)		
maxretn(Maximum daily return)			sg(Sales growth)		
sharechg(Changes in shares outstanding)			rg(Revenue growth)		
cf(Cash flows)					

Sentiment Factors

Direct sentiment factors:(from Stocktwitz, Sentdex, twitter)

1.**Sentiment** signal determined by the Sentdex algorithm. This is a continuous value ranging from -3 to 6.

2. bullish/bearish intensity: PsychSignal's algorithms score each message for the strength of bullishnes/bearishness present on a 0-4 scale.

3. bull_minus_bear: This indicator simply subtracts bearish_intensity from bullish_intensity to provide an immediate net score

4. **bull/bear_scored_messages**: The total count of bullish/bearish sentiment messages scored by the PsychSignal's algorithm

5. bull_bear_msg_ratio: Ratio between bull_scored_messages and bear_scored_messages

6. No. of total messages: The number of messages coming through

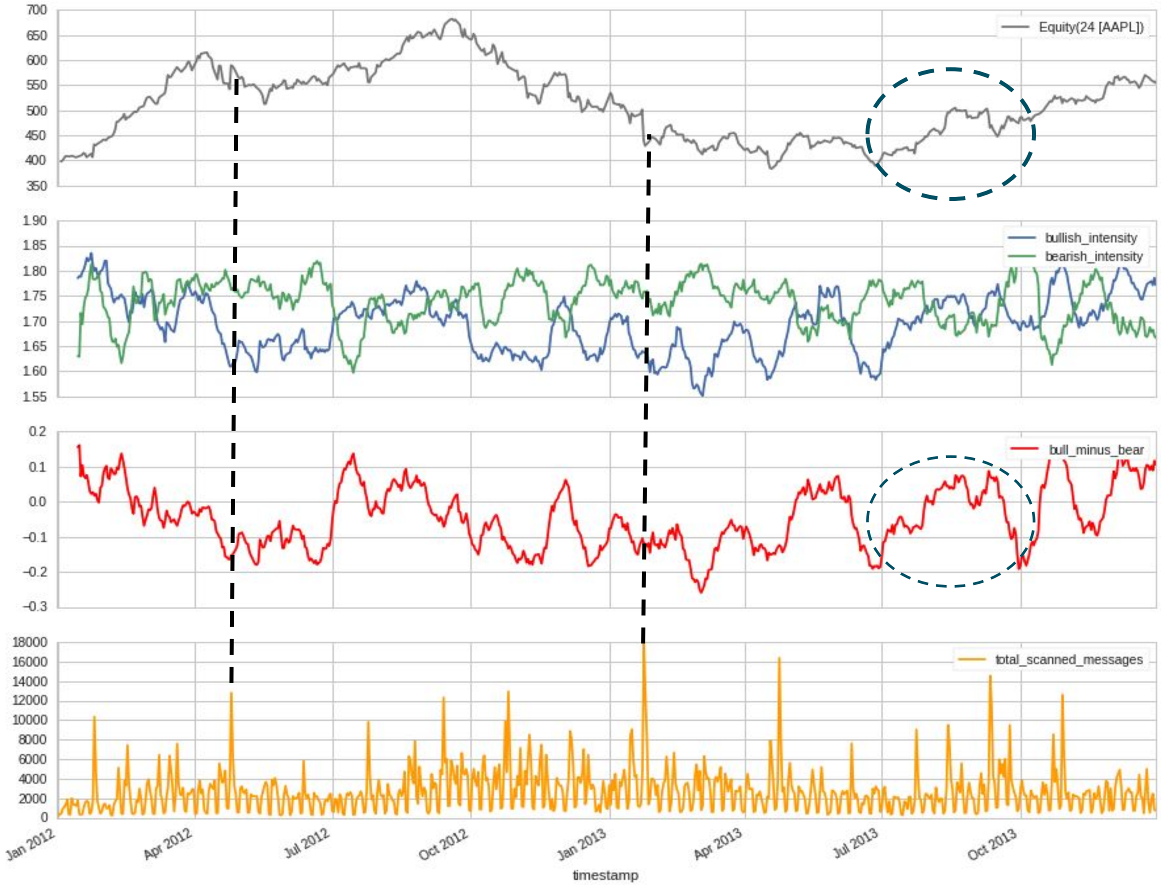
** factors are computed with different time frames using simple moving average

Indirect sentiment factors:

1. number of new IPOs
2. first day IPO return
3. closed-end fund discount
4. durables consumption index
5. nondurables consumption index
6. services consumption index
7. consumer price index
8. industrial production index
9. employment data
10. market turnover rate(liquidity)

** Other financial factors from previous part such as price-earning ratio and dividend premium

Sentiment Factors -- direct factors example



Conclusions:

> A spike in the total message volume may correlate with a drop in price.

> There are not very much difference between bullish and bearish intensities of the same time period(the difference is located in -0.2 ~ 0.2)

> correlated trends with prices and bull-minus-bear intensities

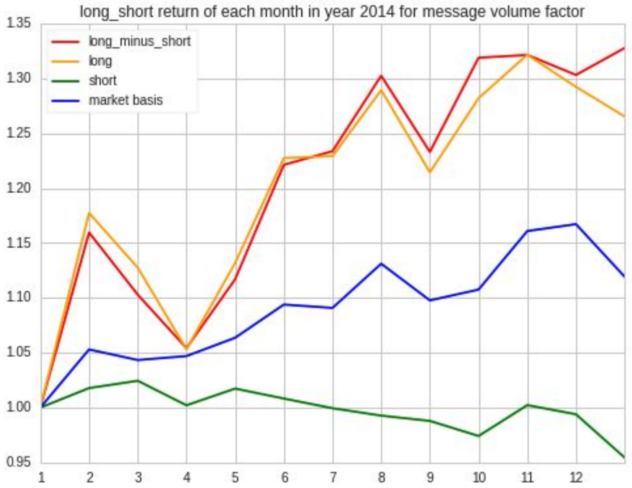
> a large increase in the bull_minus_bear intensities normally followed by an increase in the return

Factor-effectiveness in different time period(2013 to 2017)

2013		2014		2015		2016		2017	
Factors	Sharpe	Factors	Sharpe	Factors	Sharpe	Factors	Sharpe	Factors	Sharpe
Message volume	1.59	Bullish intensity_3	2.458	Bearish latest	3.053	Bearish score	2.130	Bearish score	1.991
Senti_return combined	1.308	Senti_sma_20	2.366	Bull_minus_bear_latest	2.418	Message volume	1.94	Bull_minus_bear_sma	1.781
Bullish intensity_3	1.279	Bearish score	2.245	Bullish_latest	1.564	Bullish score	1.846	Senti_return combined	1.654
Bull-bear-message ratio	1.124	Senti_sma_50	2.144	Bullish score	1.508	Bullish_latest	1.846	Bullish_latest	1.556
		Senti_sma_30	1.939	Bullish intensity_3	1.298	Senti_return combined	1.619	Bull messages	1.386
		Bull messages _3	1.915	Bull messages	1.288	Senti_sma_50	1.598	Bull messages_12	1.242
		Bull-bear-message ratio	1.828	Senti_latest	1.118	Bullish intensity_3	1.334	Combined sentiment	1.008
Combined sentiment	1.369	Bearish latest	1.738	Combined sentiment	1.075	Bull messages	1.303		
Bullish intensity_3	1.207	Message volume	1.696	Bull messages_12	1.04	Senti_sma_30	1.143		
Senti_latest	1.204	Bullish score	1.613			Bull_minus_bear_sma	1.112		
Bull messages	1.132	Bull messages_12	1.586			Combined sentiment	1.004		
Important factors	**4**	**Important factors**	**15**	**Important factors**	**9**	**Important factors**	**11**	**Important factors**	**7**

Factors frequencies

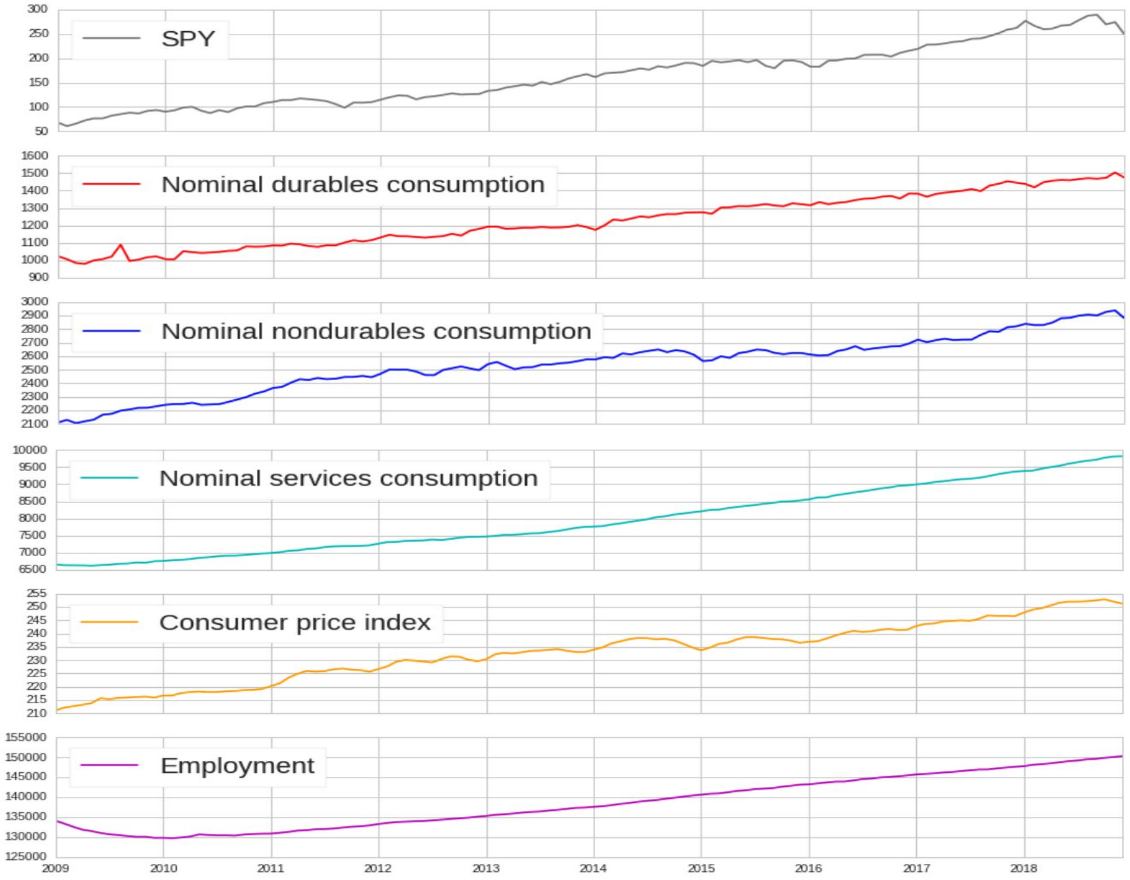
Index	Frequency	Index	Frequency
Average of 3-day bullish intensity	5	Average 3-day bull_minus_bear score	2
Combined 3-day scores from Sentdex and Stocktwitz	4	Latest bearish score	2
Bull messages	4	Average 3-day bull-bear-message ratio	2
Total messages	3	Latest Sentdex sentiment	2
Combined 12-day bearish score from Stocktwitz and twitter	3	Sentdex sentiment 30-day	2
Combined 12-day bearish score from Stocktwitz and twitter	3	Sentdex sentiment 50-day	2
Bullish latest scores from Stocktwitz	3	Sentdex sentiment 20-day	1
Average bull messages for 12-day	3	Bull_minus_bear_latest	1
Combined 5-day sentiment with returns	3	Average 3-day bull messages	1



Conclusion:

- Direct sentiment factors are correlated with the market, the message, intensity factors both seem significant and have a prediction power
- Sentiment factors are also dependent on time, some factors may be effective during one particular year while not in other years
- A few combined factors seem to be effective as well(from twitter, Stocktwitz, Sentdex)
- A time window of 3-day used to compute the average scores seems to be beneficial

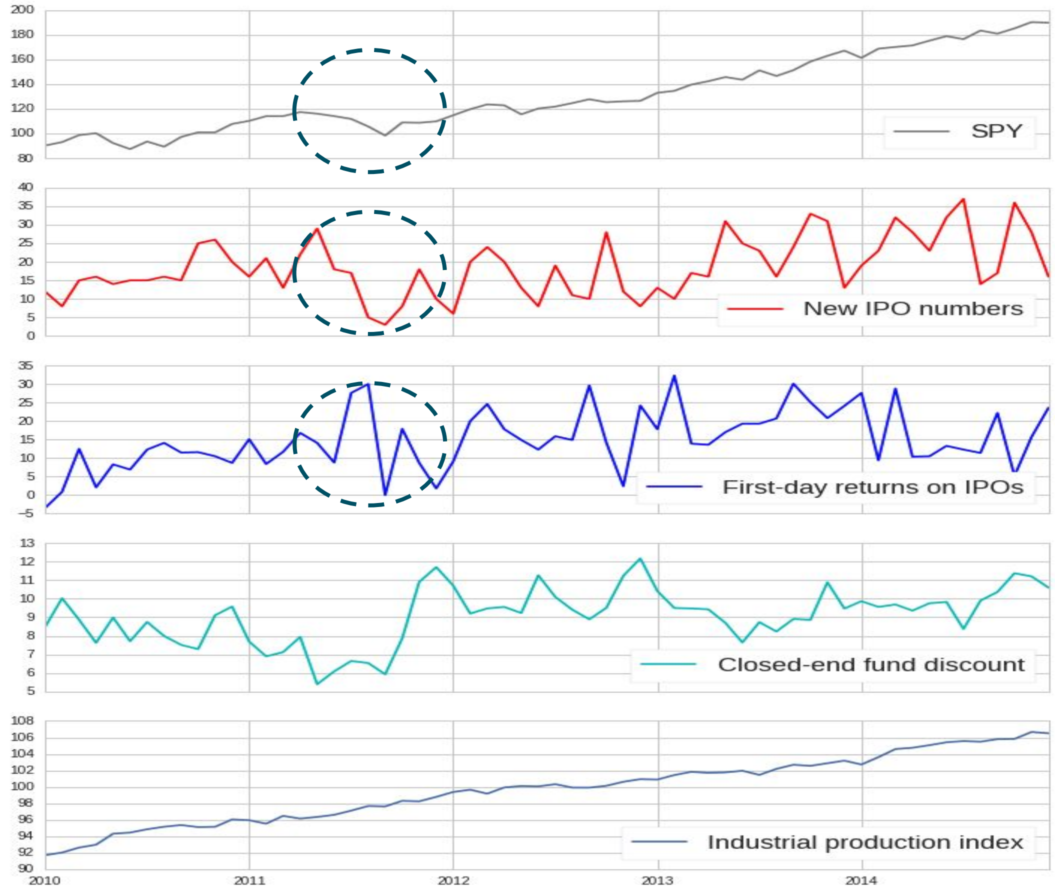
Sentiment Factors -- indirect factors example



-- For the 10-year period from 2009 to 2018, the following five indices seem to have a good prediction power of the returns:

1. durables consumption index
2. nondurables consumption index
3. services consumption index
4. consumer price index
5. employment data

Sentiment Factors -- indirect factors example



<u>For the five-year period from 2010 to 2015:</u>

1. The industrial production index seems to have a good prediction power
2. For some specific data period, an increase in the number of new IPOs, first day returns on new IPOs and closed-end fund discount followed by an increase in returns, a huge drop followed by a decrease in SPY prices
3. These factors may not be that significant compared to the previous slides

Factor Analysis

- Six classes, 38 normal factors
- + Sentimental factors, 19 direct, 12 indirect
- Factor Analysis through 2009 to 2017, nine years
- **Take year 2017 as an example**
  - We selector 2018 prediction factors based on factors performance in 2017
  - We use four indexes to valuate factors, *annualized_return, annualized_volatility, sharp ratio, max drawdown*

$$sharp\ ratio = \frac{R_p - R_f}{\sigma_p}$$

Max draw down the maximum observed loss from a peak to a trough of a portfolio, before a new peak is attained. Maximum drawdown is an indicator of downside risk over a specified time period

$$Max_{drawdown} = {}^{though\ value - \text{peak value}}/_{Peak\ value}$$

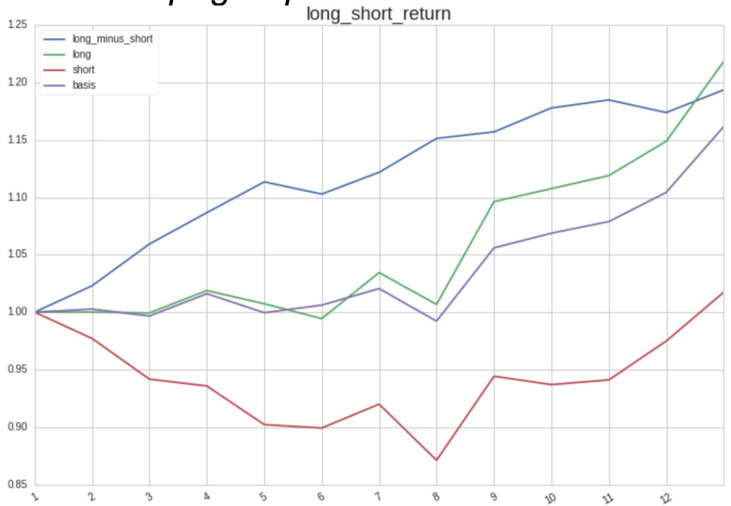
Factor Analysis

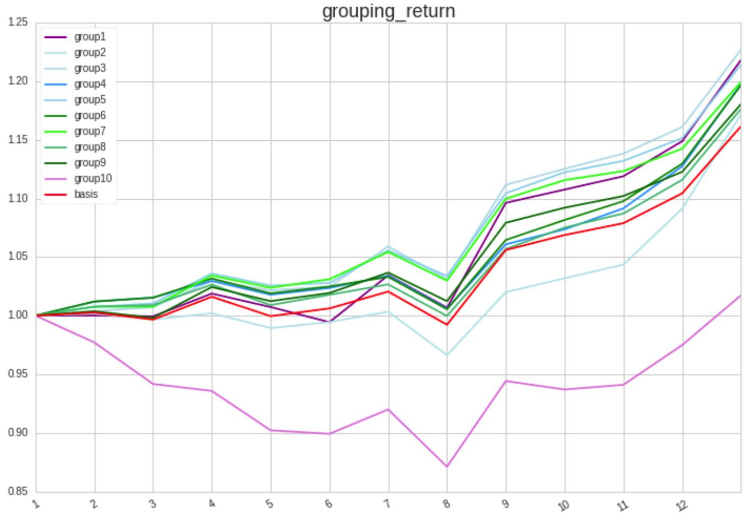
- **Take year 2017 as an example (continued)**
  - We compute the four measurement for our 38 computed factors, with 20 sentiment factors
  - *We long the top 10% stocks, and short bottom 10% stocks in the first trading day each month, then compute factor return*(to lower the risk)
  - Assume that free-risk rate is 2%
  - We sort factors by its Sharpe ratio, the key valuation index
  - We get 20 basic factors with top 20 Sharpe ratio as the prediction factor for 2018, and top 10 sentiment factors
  - There are 7 basic factors' Sharpe ratio > 2
  - 17 basic factors' Sharpe ratio > 1, and 7 sentiment factors' Sharpe ratio > 1
  - The *oig(operating income growth)* has top performance in terms of Sharpe ratio 3.69, with return 19.35%, volatility 4.71%
  - Mean of max_drawdown is 8.73%, median 8.51%, all below 15%

		annualized return	annualized volatility	sharpe ratio	max_drawdown
Operating income growth	oig	0.193533	0.047072	3.686569	0.043564
Revenue growth	rg	0.324957	0.094977	3.210854	0.093339
Sales growth	sg	0.246889	0.083456	2.718671	0.082583
Changes in shares outstanding	sharechg	0.170140	0.064512	2.327301	0.061537
PEG ratio	peg	0.160150	0.060752	2.306928	0.057520
EPS growth	epsg	0.082459	0.029796	2.096240	0.032018
Return over assets	roa	0.213041	0.095836	2.014285	0.084363
Return over equity	roe	0.189501	0.085986	1.971258	0.081629
Maximum daily return	maxretn	0.268326	0.134365	1.848144	0.108884
Net income growth	nig	0.100179	0.044602	1.797649	0.045268
Quick ratio	qr	0.102355	0.049452	1.665346	0.034267
EV to EBITDA	ev_ebitda	0.126239	0.072197	1.471514	0.066291
Total assets growth	ag	0.136120	0.087147	1.332468	0.080595
Turnover	turn	0.124588	0.084796	1.233398	0.089538
Current ratio	cr	0.091463	0.062172	1.149437	0.042726
Cash flow to price ratio	cfp	0.144183	0.118451	1.048397	0.074747
Asset Turnover	ato	0.114848	0.091440	1.037265	0.085767

Factor Analysis

- **Take year 2017 as an example (continued)**
  - We select *oig* as an example of factor analysis
    - *Oig is the operating income growth in quarterly reports*
    - Firstly, we plot the return of *oig* factor in 2017
      - ✔ Long group has good performance, 21.82%, compared with basis 16.15%
      - ✔ Short group only has 1.73% return, good differentiation
    - Then we plot the return of 10 groups differentiated by factor. Factor *can differentiate groups well, except group5*





Factor Analysis

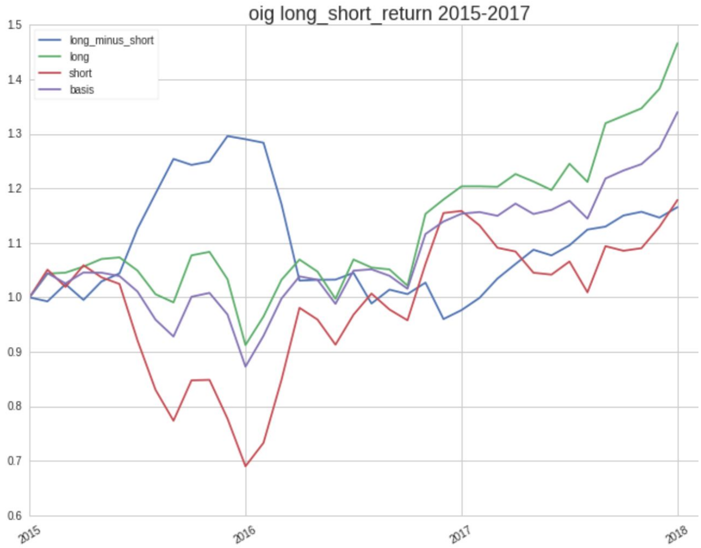
1.Market
2.Momentum
3.Value
4.Growth
5.Profitability
6.Liquidity

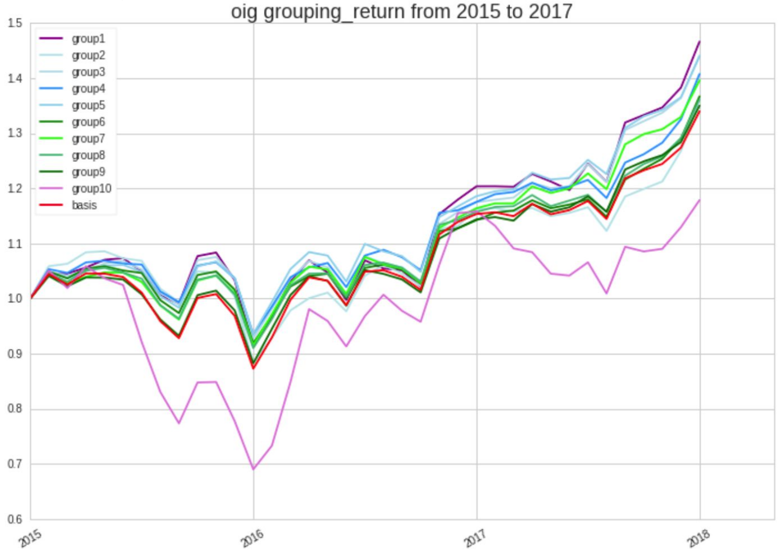
- We repeat formal steps each year from 2009 to 2017, get following result
- **Frequency statistics for factors** (top 20 sharpe ratio each year from 2009 to 2017)
  - No factor can occur every year, each year change
  - All classes have factors with good performance

Index	Frequency	Index	Frequency	Index	Frequency
nig	8	mom36	5	skew	4
vol	8	sg	5	std_turn	3
bp	7	betasq	5	cf_sale	3
sharechg	7	ato	5	cr	3
mom12	7	roe	5	cf	3
size	6	roa	5	volumed	3
maxretn	6	beta	5	rg	3
epsg	6	dpsg	5	ev_ebitda	3
nig	6	bpsg	4	largretn	3
cfp	6	qr	4	peg	2
sp	6	ag	4	momchg	2
fcf	6	turn	4	ep	1
mom6	6	std_volumned	4		

Factor Analysis

- Oig Factor analysis through 3 years
  - From 2015-01-01 to 2017-12-31
  - Long group has good performance, 13.61%, higher than compare basis 10.24%
  - Short group only has 5.63% return, good differentiation
  - Long and short 10% stock can achieve annualized return of 5.24%
  - Still has good differentiation, especially the group 10
  - Factor’s performance in short period may better than longer period





Factor Analysis

- Factor analysis through 3 years
  - We did the same factor analysis with the period 2015-01-01 to 2017-12-31
  - The factor return is relatively low because factor effectiveness changes each year
  - ***Therefore, in our model we re-select our effective factors in the first month every year***

	annualized return	annualized volatility	sharpe ratio	max_drawdown
qr	0.059021	0.072818	0.535866	0.095738
fcf	0.068379	0.090907	0.532184	0.107813
cf_sale	0.079116	0.117348	0.503766	0.168466
sharechg	0.054280	0.082737	0.414329	0.113644
ato	0.065946	0.138383	0.332019	0.159320
bpsg	0.048453	0.088065	0.323094	0.095738
ep	0.049834	0.093012	0.320753	0.101733
cfp	0.069245	0.176185	0.279511	0.197651
rg	0.060415	0.150469	0.268596	0.180635
oig	0.052445	0.131491	0.246746	0.183852
sg	0.054703	0.143238	0.242276	0.170606
peg	0.041391	0.107121	0.199694	0.120892
ag	0.045382	0.129910	0.195380	0.153302
volumned	0.040540	0.118327	0.173583	0.155849
roa	0.046543	0.153461	0.172966	0.225822
std_volumned	0.039040	0.114573	0.166178	0.153490
size	0.047321	0.166024	0.164563	0.232344

Prediction models

- **Models**: we use 6 different linear models as our prediction model
  - OLS, Ridge, Bayesian ridge, Lasso, Elastic Net, PLS
  - Also, we consider a forecasting combination model(FC) based on these 6 linear models

- **Methods**:
  - Our prediction is built on a monthly rolling basis
  - For each year, the input factor is determined by the factor analysis of previous year
  - Then for each month of this year,
    - **Input**: factor data of last 12 month
    - **Output**: return of this month
  - Example:
    - Input data of 2010-02 is the 2009's significant factors from 2009-02 to 2010-01
    - Input data of 2013-07 is the 2012's significant factors from 2012-07 to 2013-06

Prediction models

To test the performance of our prediction methods and the effectiveness of sentiment factor, our prediction includes two parts:

- **1. Prediction without sentiment data:**
  - Prediction window: 9 years, from 2010-01-01 to 2018-12-31
  - Input: top 20 factors (out of 38)

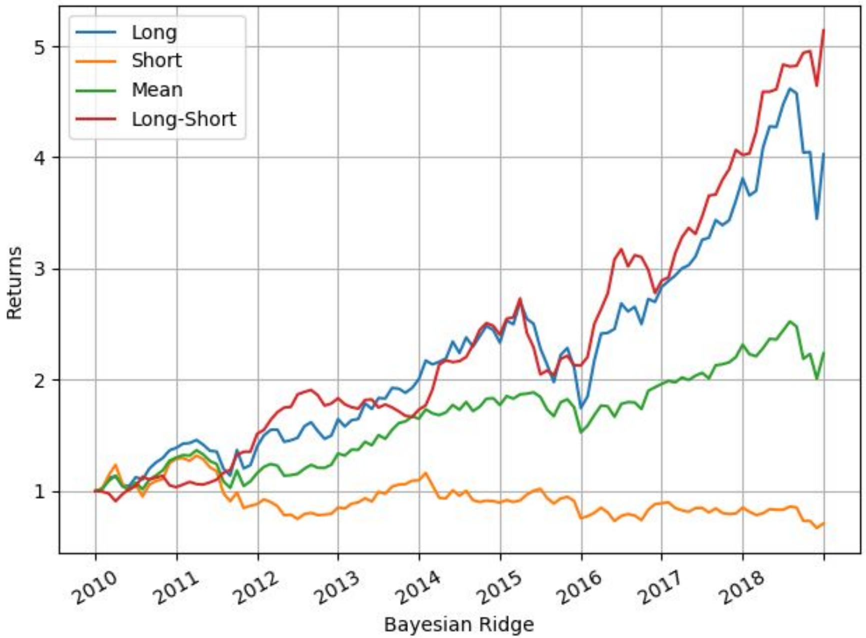
- **2. Prediction with sentiment data:**
  - Prediction window: 5 years, from 2014-01-01 to 2018-12-31
  - Input: top 20 factors (out of 38), top 10 sentiment factors (out of 19), 12 indirect sentiment factors

- **Data preprocessing**:
  - There are about valid 1000 stocks each month, input shape: (**12 * num of stocks**) * **num of factors**
  - We scale the input factor data by: $X_{scale} = \frac{X - \bar{X}}{\sigma_X}$

- **Portfolio strategies**:
  - Long 5% of stocks with highest prediction return
  - Short 5% of stocks with lowest prediction return

Performance of prediction methods

Without sentiment data, from 2010 to 2018

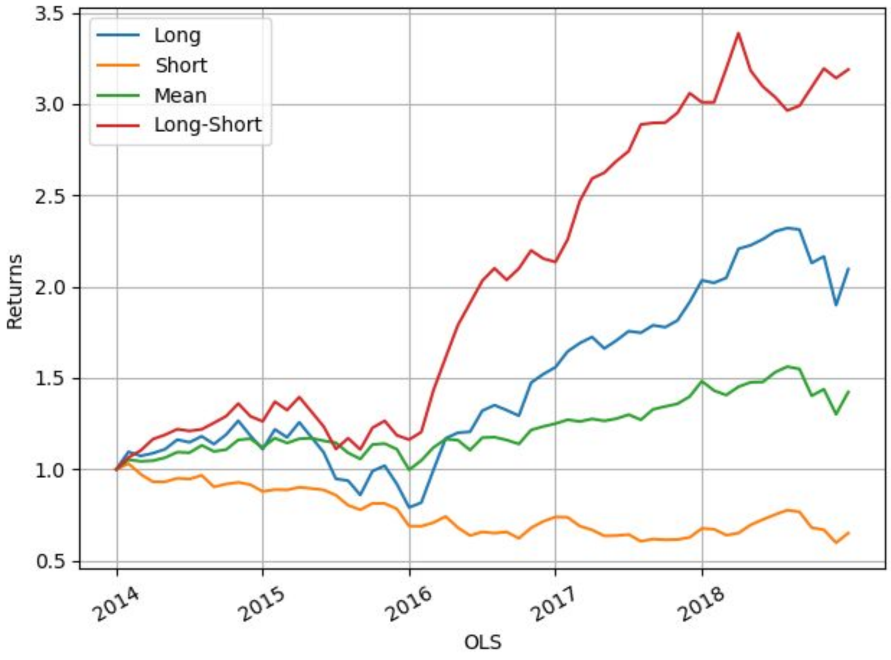


	Annualized Return	Annualized Volatility	Sharpe Ratio	Max Drawdown
Bayesian Ridge	0.1995	0.1564	1.1477	0.2082
Ridge	0.1874	0.1481	1.1305	0.2075
OLS	0.1867	0.1481	1.1252	0.2075
FC	0.1859	0.1563	1.0615	0.2147
ElasticNet	0.1812	0.1615	0.9983	0.2187
Lasso	0.1773	0.1628	0.9659	0.2361
PLS	0.1758	0.1618	0.9627	0.2147

Performance of sentiment factors

Without sentiment data, from 2014 to 2018

With sentiment data, from 2014 to 2018



Statistics of long-short portfolio

Without sentiment data, from 2014 to 2018

	Annualized Return	Annualized Volatility	Sharpe Ratio	Max Drawdown
OLS	0.2557	0.1836	1.2832	0.2254
Lasso	0.2744	0.2086	1.2194	0.3144
Ridge	0.2456	0.1891	1.1930	0.2597
FC	0.2457	0.1946	1.1602	0.2401
ElasticNet	0.2383	0.2016	1.0827	0.2826
Bayesian Ridge	0.2154	0.2005	0.9749	0.2611
PLS	0.1839	0.1872	0.8756	0.2313

With sentiment data, from 2014 to 2018

	Annualized Return	Annualized Volatility	Sharpe Ratio	Max Drawdown
ElasticNet	0.2775	0.1861	1.3836	0.2119
OLS	0.2611	0.1778	1.3557	0.2111
Lasso	0.2883	0.1983	1.3534	0.2409
FC	0.2711	0.1893	1.3263	0.2288
Ridge	0.2648	0.1864	1.3137	0.2338
Bayesian Ridge	0.2458	0.1952	1.1565	0.2533
PLS	0.2087	0.2024	0.9323	0.2468

Statistics of long-short portfolio with transaction cost

- To simulate the real world setting, we set a one side transaction cost of 0.05%

Without sentiment data, from 2014 to 2018

	Annualized Return	Annualized Volatility	Sharpe Ratio	Max Drawdown
OLS	0.1836	0.1836	0.8909	0.2263
Lasso	0.2013	0.2086	0.8692	0.3157
Ridge	0.1740	0.1891	0.8147	0.2608
FC	0.1742	0.1946	0.7925	0.2411
ElasticNet	0.1672	0.2016	0.7299	0.2837
Bayesian Ridge	0.1455	0.2005	0.6259	0.2622
PLS	0.1157	0.1872	0.5110	0.2322

With sentiment data, from 2014 to 2018

	Annualized Return	Annualized Volatility	Sharpe Ratio	Max Drawdown
ElasticNet	0.2043	0.1861	0.9903	0.2128
Lasso	0.2146	0.1983	0.9812	0.2419
OLS	0.1887	0.1778	0.9489	0.2120
FC	0.1982	0.1893	0.9414	0.2297
Ridge	0.1923	0.1864	0.9245	0.2347
Bayesian Ridge	0.1742	0.1952	0.7900	0.2543
PLS	0.1391	0.2024	0.5885	0.2479

Next Steps

1. Apply machine learning / deep learning networks such as RNN or TCN(temporal convolutional networks) (currently not supported on Quantopian)
2. Apply NLP techniques to compute more sentiment signals based on lexicons/bi-grams/n-grams
3. Use more measurements to evaluate factors and try to create some compound factors with better performance

References

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
5. Dhaoui, Abderrazak, and Nesrine Bensalah. "Asset valuation impact of investor sentiment: A revised Fama–French five-factor model." *Journal of Asset Management* 18.1 (2017): 16-28.
6. Chung, San-Lin, Chi-Hsiou Hung, and Chung-Ying Yeh. "When does investor sentiment predict stock returns?." *Journal of Empirical Finance* 19.2 (2012): 217-240.
7. Kim, Soon-Ho, and Dongcheol Kim. "Investor sentiment from internet message postings and the predictability of stock returns." *Journal of Economic Behavior & Organization* 107 (2014): 708-729.
8. Baker, Malcolm, and Jeffrey Wurgler. "Investor sentiment and the cross-section of stock returns." *The journal of Finance* 61.4 (2006): 1645-1680.
9. Berger, Dave, and H. J. Turtle. "Cross-sectional performance and investor sentiment in a multiple risk factor model." *Journal of Banking & Finance* 36.4 (2012): 1107-1121.
10. Changsheng, Hu, and Wang Yongfeng. "Investor sentiment and assets valuation." *Systems Engineering Procedia* 3 (2012): 166-171.

Thanks!
Q&A

---- Final Presentation

Group 4:
Yu Gu
Ruoyu Han
Yuzhu Zhang