



UCLA Applied Finance Project 2020

Sector Misclassification

Table of Contents

- Background and Data
 - Comment on ideas behind the strategy
 - Describe data and compare to results within “Categorization Bias in the Stock Market” by Kruger *et al.*
 - Report results from replication
- Strategy Reformulation
 - Explain changes made to main strategy
 - Results in- and out-of-sample
- Robustness Tests
 - Explain further variations
 - Report results in-sample
- Comments from Eric
- Conclusion
- Appendix

Background and Data

Background

- The investment strategies we are considering are based on the paper “Categorization Bias in the Stock Market” by Kruger *et al.*
- Some investors perceive individual firms through coarse industry classification. Such categorical thinking generates mispricing and predictability in stock-returns.
- For example, think of a company within the paper industry which produces cigarette paper and therefore depends on cigarette sales. If investors were fully rational, this company would react little to the returns of the other firms in the paper industry and more to its “fundamental peers” within the cigarette business. However, some investors may base their investment decisions on the performance of the paper industry, thus generating mispricing.

Formalization of Strategy

The strategy takes a long- or short-position in the stock of firm j based on the discrepancy between the returns of firm j 's official industry $r_j^{(o)}$ and its fundamental peers $r_j^{(f)}$.

The premise behind the strategy is as follows:

- The divergence between firm j 's industry returns $r_j^{(o)}$ and its fundamental peer returns $r_j^{(f)}$ implies mispricing of firm j 's stock.
- The fundamental returns $r_j^{(f)}$ gives a better estimate of how firm j 's stock returns ought to behave than the returns of its industry $r_j^{(o)}$.

Fundamental Peers

- The fundamental peers of firm j are determined using **similarity scores generated by Hoberg and Phillips**, which we obtained from their website.
- The similarity scores were calculated using firms' **product descriptions within their 10-K filings** on the EDGAR database on the SEC website.
- The **rarest 25% of nouns and proper nouns** form a dictionary of words. The dictionary is ordered.
- Each firm has a word-vector of words included in their product description. The word-vectors are normalized. Raw similarity scores are obtained by taking the dot product of firms' normalized word-vectors to obtain the **cosine similarity**.
- Hoberg and Phillips similarity scores **truncate** and **normalize** the raw scores. **The similarity scores are between 0 and 1.**

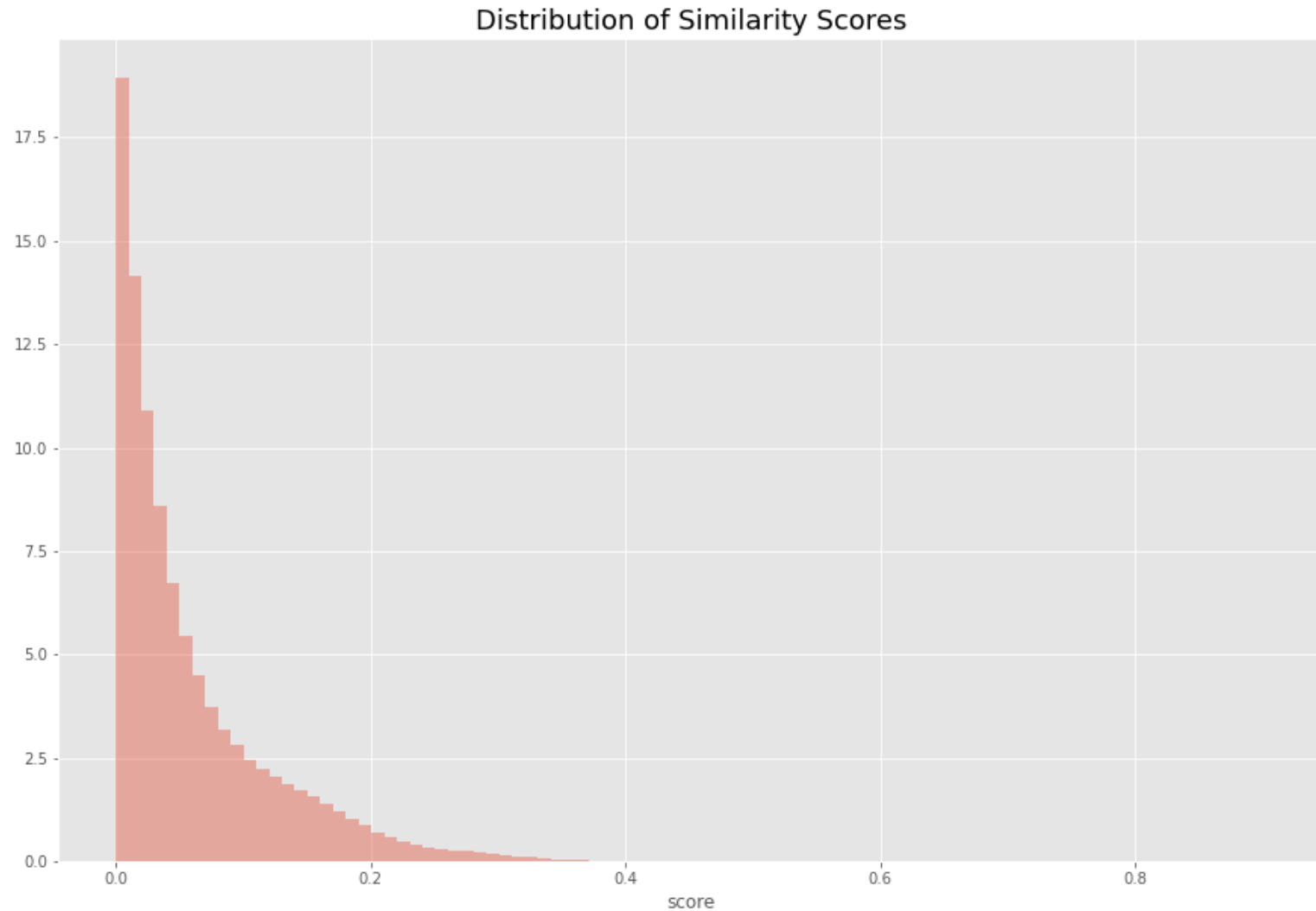
Example of Fundamental Peers

To illustrate the concept of fundamental peers, consider the example of **Morgan Stanley** in December 2013. Morgan Stanley's industry was **Capital Markets**.

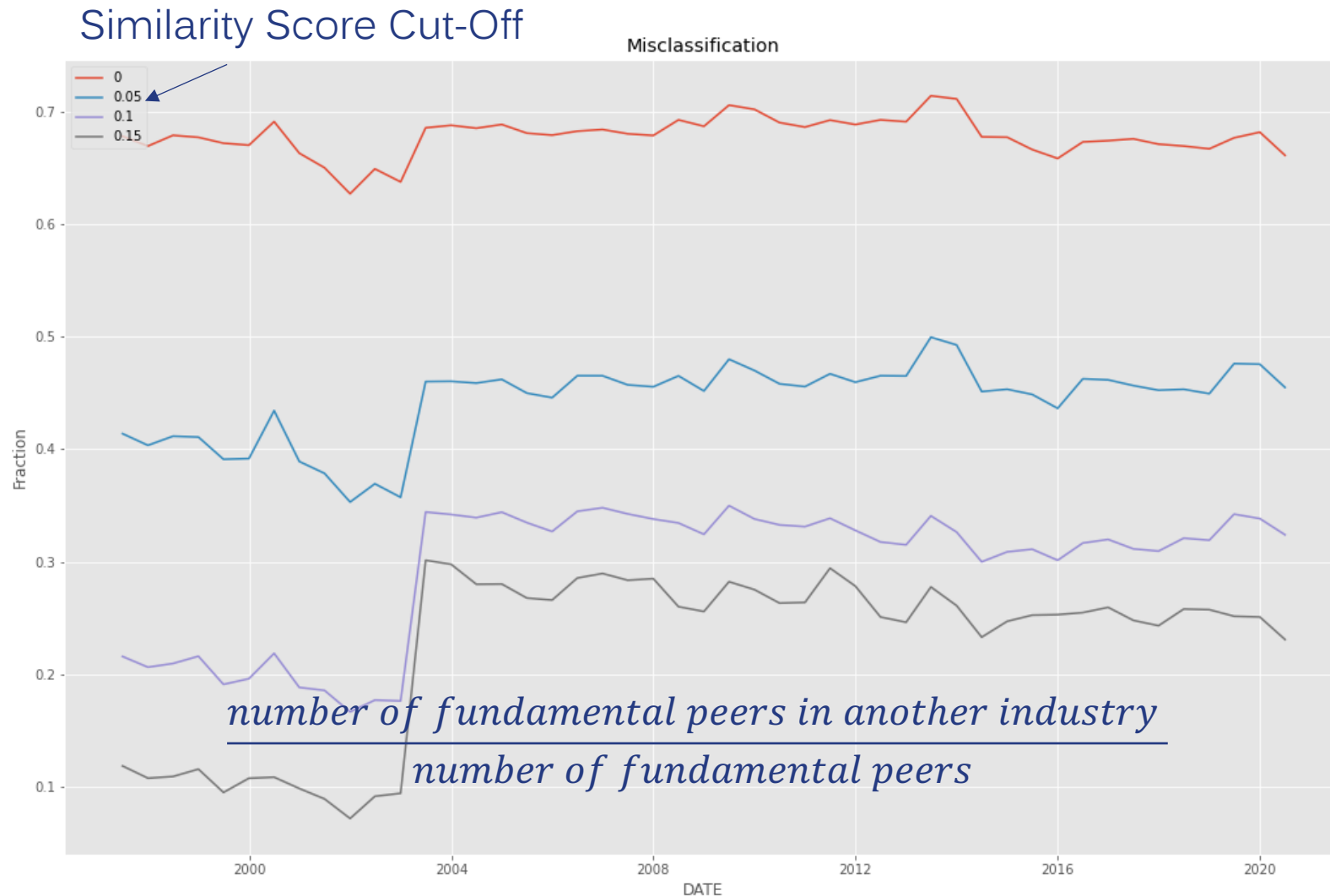
- **Goldman Sachs** and **BlackRock** were fundamental peers and are also in the industry **Capital Markets**.
- **JPMorgan Chase** was a fundamental peer but is in the industry **Diversified Financial Services**.
- **AIG** was a fundamental peer but is in the industry **Insurance**.

Using similarity scores above 0.05, Morgan Stanley's fundamental peers contained 107 members of which 79% were from industries other than Capital Markets.

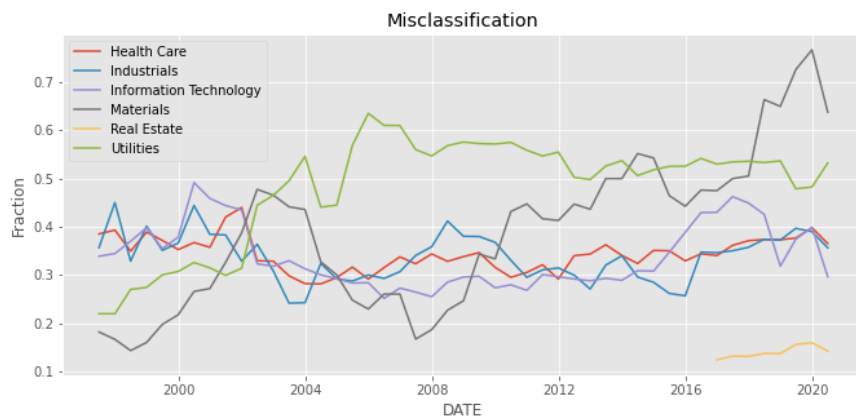
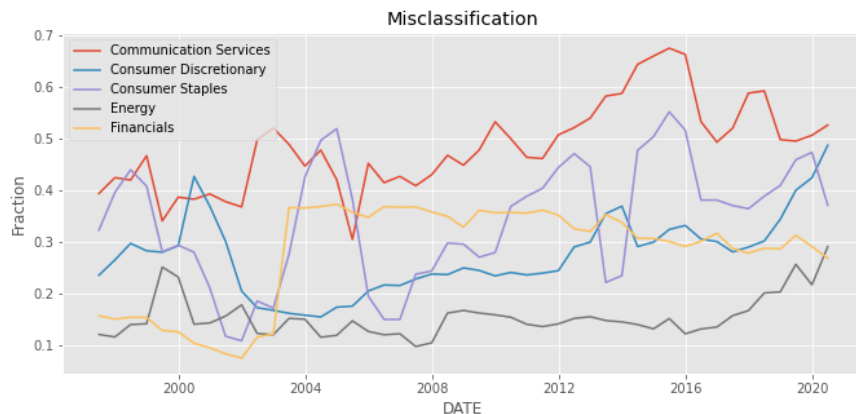
Distribution of Similarity Scores



Misclassification Ratio



Misclassification Ratio by Sector



Sector	Mean	SD
Real Estate	25.4%	1.8%
Energy	33.8%	3.7%
Financials	39.7%	8.2%
Consumer Discretionary	40.6%	5.7%
Health Care	47.2%	2.8%
Consumer Staples	52.1%	8.4%
Information Technology	52.6%	3.8%
Materials	55.5%	9.7%
Industrials	61.1%	5.4%
Utilities	64.8%	7.9%
Communication Services	66.4%	4.2%

Data Description

	Kruger	AFP
Return Data	CRSP US (monthly and weekly) 1997-2009	MSCI IMI US (monthly) 1997-November 2020
Official Industry	SIC2 (Compustat)	GICS industry (MSCI)
Similarity scores	Hoberg and Phillips Data (annual)	Hoberg and Phillips Data (annual) 1989-2019

- Year y 's Hoberg and Phillips data is used in year $y + 1$.

Subsetting (from Kruger):

- Remove all firms with fewer than 5 industry or fundamental peers.
- Remove firms whose share price is less than \$5.
- Remove firms with negative or 0 book equity.
- We only included firms with a Hoberg and Phillips similarity score greater than 0.05 so that the size of the groups would match the paper's.

Comparison of Data 1997-2009

AFP	Mean	SD	P25	Median	P75
Firms in GICS industry	103.886	70.833	57.000	91.000	132.000
Fundamental peers	77.835	88.474	15.000	43.000	102.000
$\ln(M)^*$	13.955	1.488	12.884	13.749	14.828

Kruger	Mean	SD	P25	Median	P75
SIC2 firms	224.786	188.681	66.000	140.000	372.000
Fundamental peers	83.278	34.000	118.682	34.000	98.000
$\ln(M)^*$	12.535	2.156	10.942	12.472	13.989

* Market equity measured in thousands and is calculated at a monthly frequency

Comparison of Data 1997-2009

AFP	Mean	SD	P25	Median	P75
r_j	0.013	0.148	-0.059	0.008	0.077
$r_j^{(o)}$	0.009	0.086	-0.031	0.011	0.052
$r_j^{(f)}$	0.013	0.101	-0.033	0.0133	0.058
$r_j^{(o)} - r_j^{(f)}$	-0.004	0.051	-0.019	-0.001	0.014

Kruger	Mean	SD	P25	Median	P75
r_j	0.012	0.201	-0.078	0.000	0.079
$r_j^{(o)}$	0.011	0.084	-0.035	0.012	0.053
$r_j^{(f)}$	0.012	0.104	-0.039	0.010	0.057
$r_j^{(o)} - r_j^{(f)}$	-0.001	0.068	-0.024	0.001	0.025

Kruger Strategy

- Equal-weighting for the calculations of industry returns $r_j^{(o)}$ and fundamental returns $r_j^{(f)}$.
- Use the value of $r_j^{(o)} - r_j^{(f)}$ at time $t - 1$ as the signal at time t .
- Place firms into quintiles based on their signal—Q1 corresponds to the smallest values of the signal and Q5 corresponds to the largest values.
- Long Q1 and short Q5.

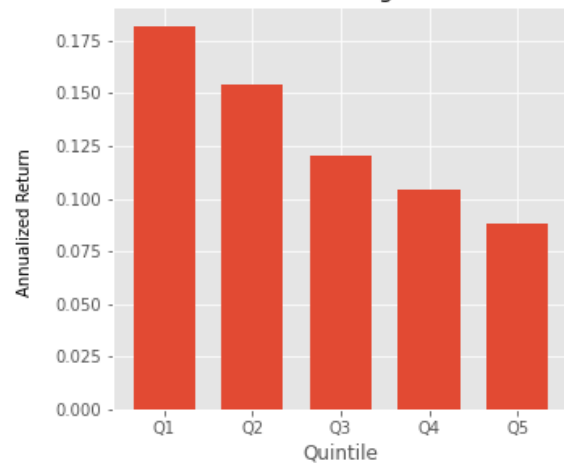
Kruger Strategy 1997-2009



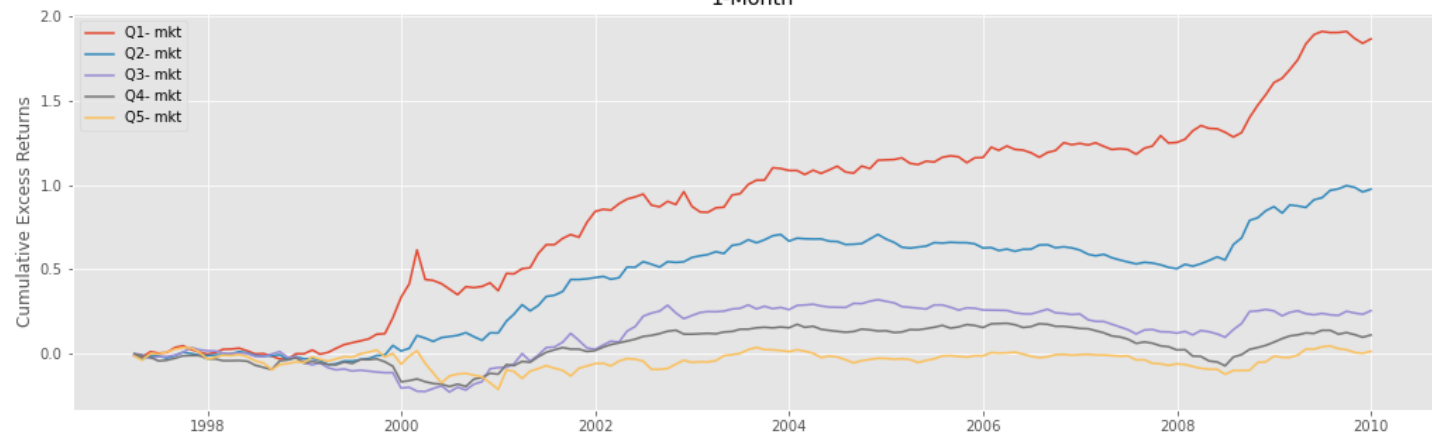
Kruger Strategy 1997-2009

	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe
Q1	19.7%	23.6%	0.70	0.83
Q5	11.5%	24.6%	0.34	0.47
Q1-Q5	8.2%	9.3%	0.53	0.88

1-Month Signal



Cumulative Returns in Excess of the Market
1-Month

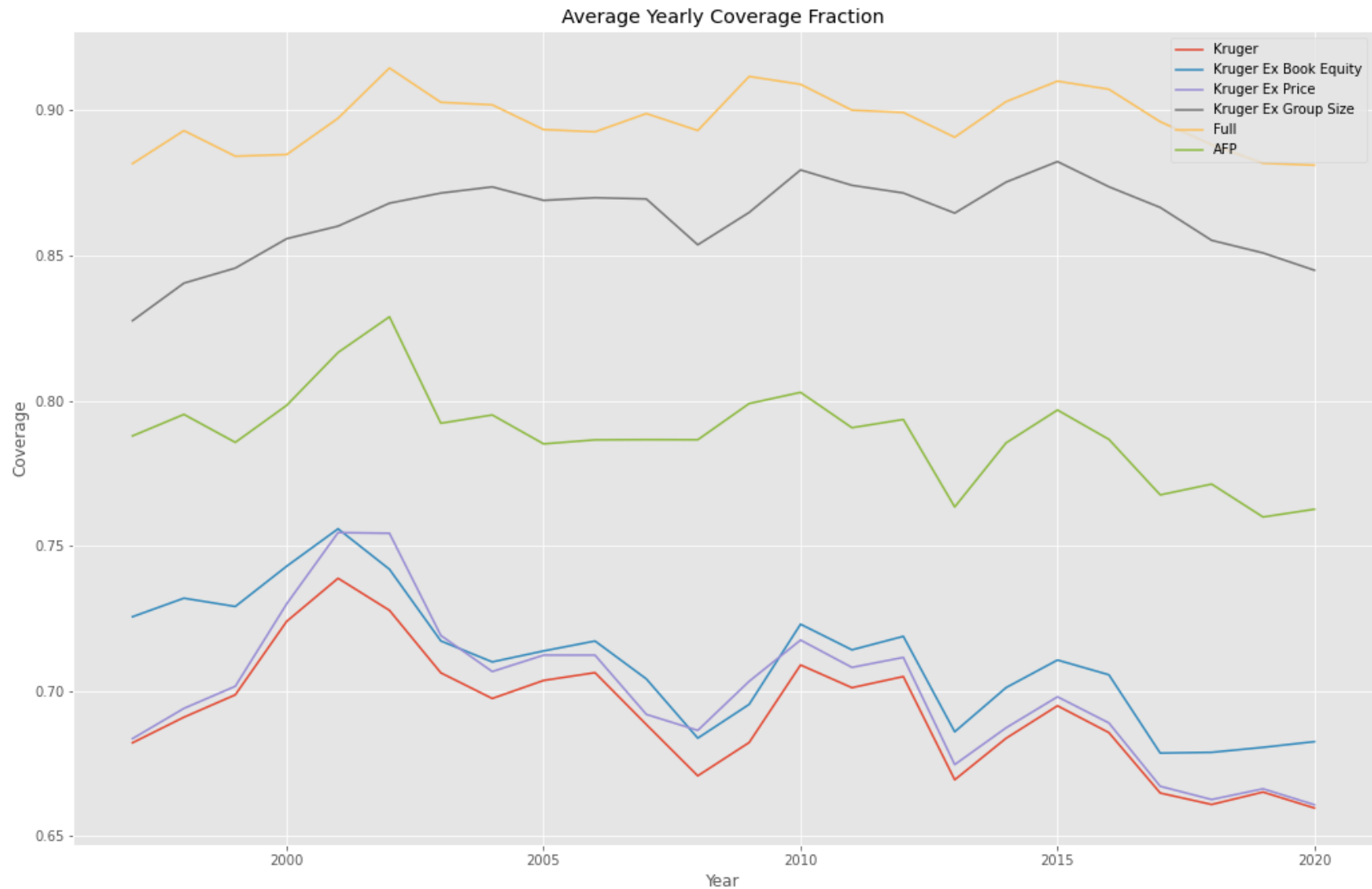


Strategy Reformulation

Strategy Reformulation

- Subset observations to firms in the monthly SIG US universe.
- Change time range to 1997-2020.
- Data from 1997-2018 was in-sample and data after December 2018 is out-of-sample.
- Expand coverage by relaxing subsetting:
 - Remove the \$5 minimum stock price constraint.
 - Remove the positive book equity constraint.
 - Reduce the minimum group size to three—this expands coverage the most.

Coverage (One-Month Signal)

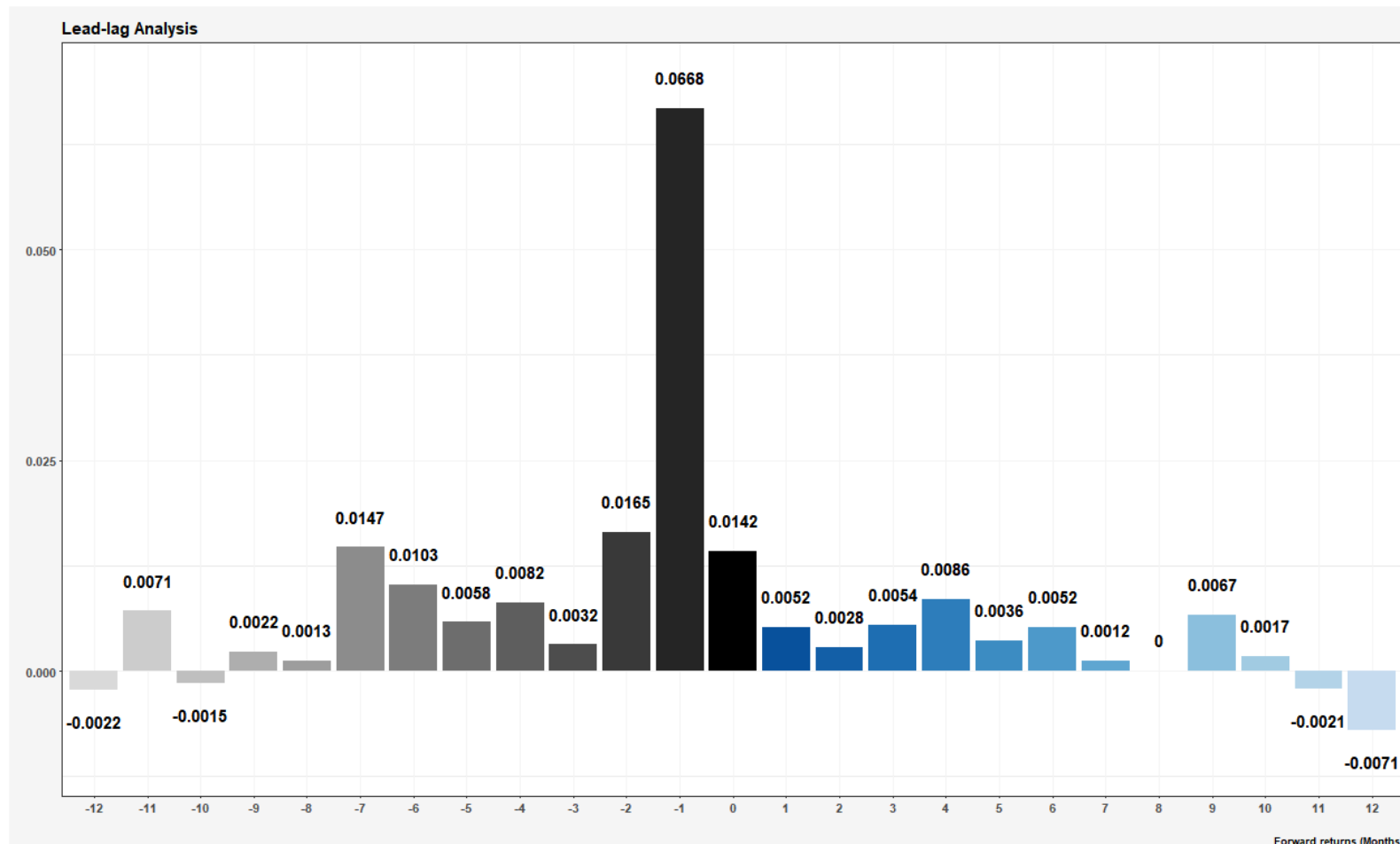


Signal and Strategy

- We will examine two signals within this section:
 - One-month signal: value of $r_j^{(o)} - r_j^{(f)}$ at time $t - 1$.
 - Six-month signal: cumulative returns of longing the official industry group and shorting the fundamental peer group from time $t - 6$ to $t - 1$
- Equal-weighting for industry and fundamental return calculations.
- The signal computation still uses the MSCI IMI US universe.
- Long Q1 and short Q5 and score-weighted portfolio.
- Portfolio weights of long- and short-leg always add to 1 and -1, respectively.

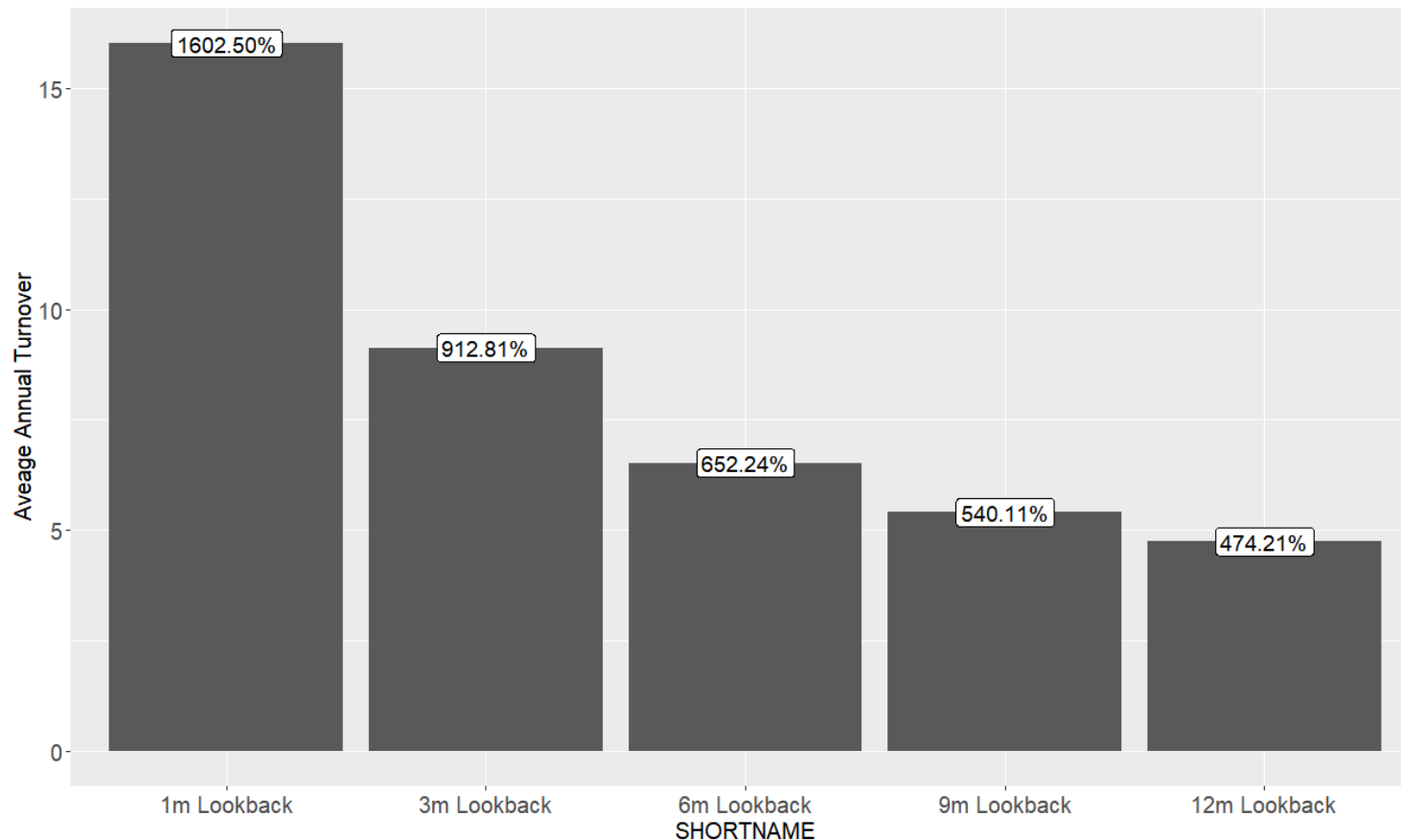
Lead-Lag ICs

- IC is 0.0052 for the first month
- The signal still appears to have predictive power going to the 10th month before showing negative IC
- There is clearly a strong one-month momentum profile



Signal Duration Comparison - Turnover

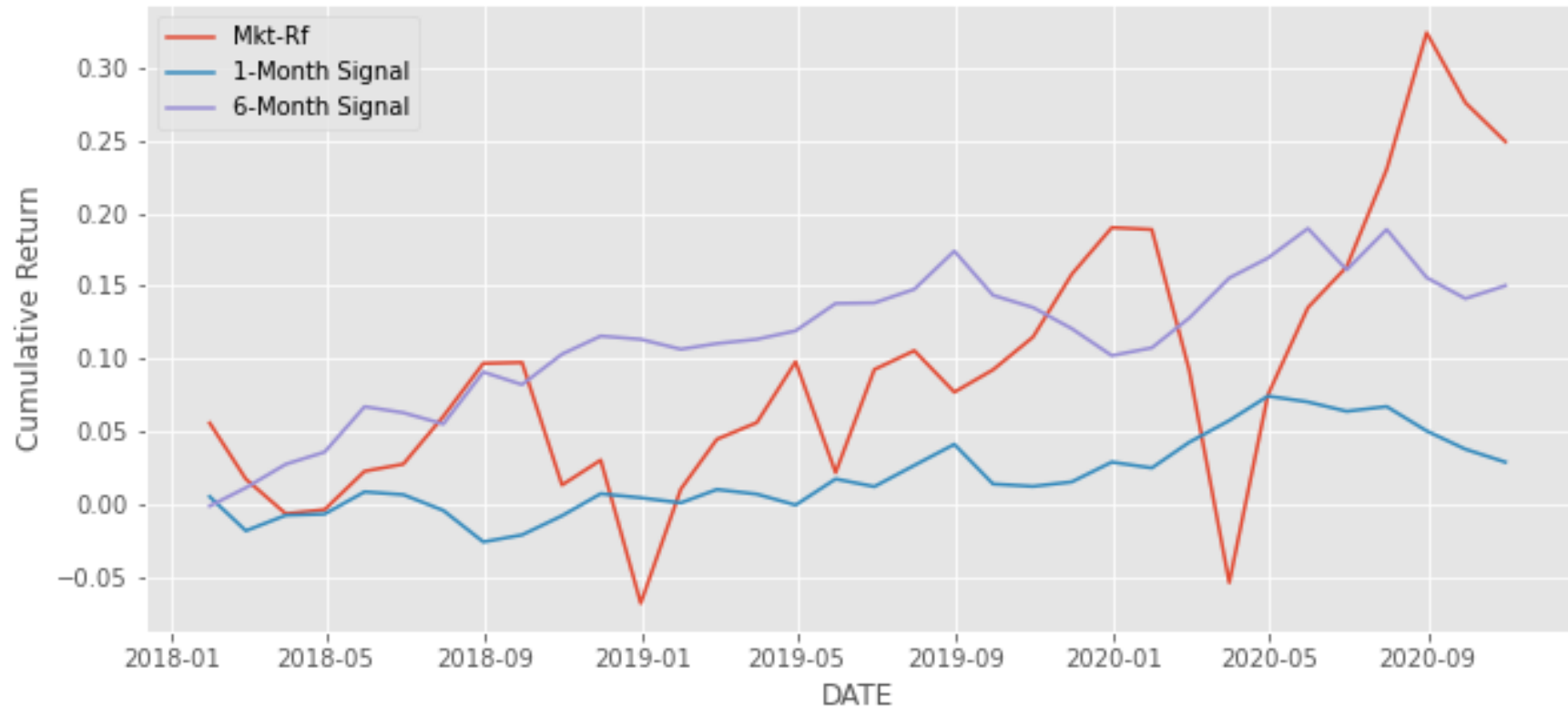
- The two-way turnover of score-weight portfolios using 1-month is about 1600% annually and is about 650% for the 6-month look back signal.



Q1-Q5 Returns 1997-2018



Q1-Q5 Returns 2019-2020 (Out-of-Sample)



Q1-Q5 Return Statistics (Annualized)

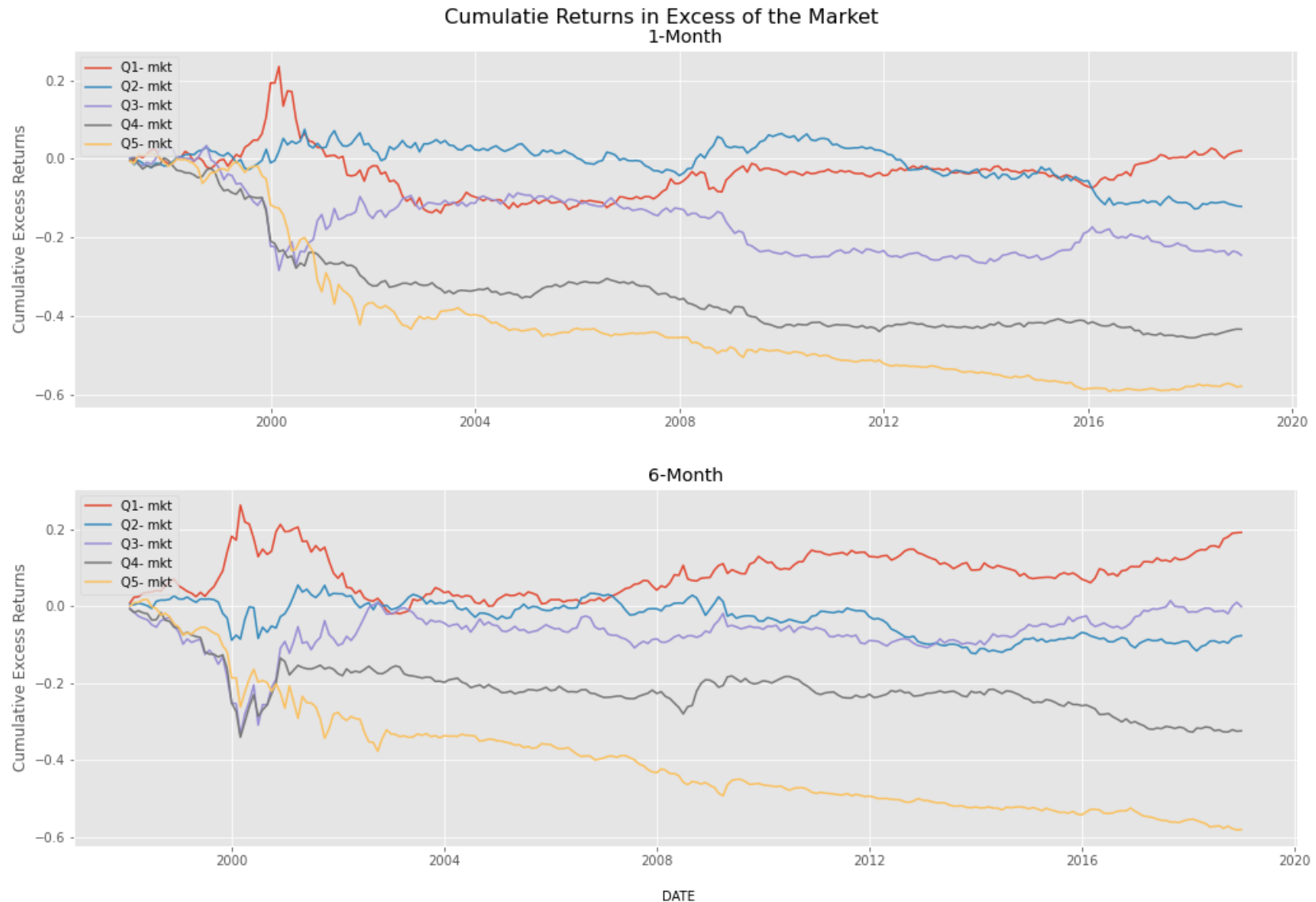
One-Month Signal

1997-2018	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe	2019-2020	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe
Q1	11.9%	19.4%	0.51	0.61	Q1	13.0%	23.5%	0.28	0.55
Q5	7.9%	21.1%	0.28	0.37	Q5	12.4%	24.6%	0.22	0.51
Q1-Q5	4.0%	8.3%	0.24	0.48	Q1-Q5	0.6%	4.3%	-0.10	0.14

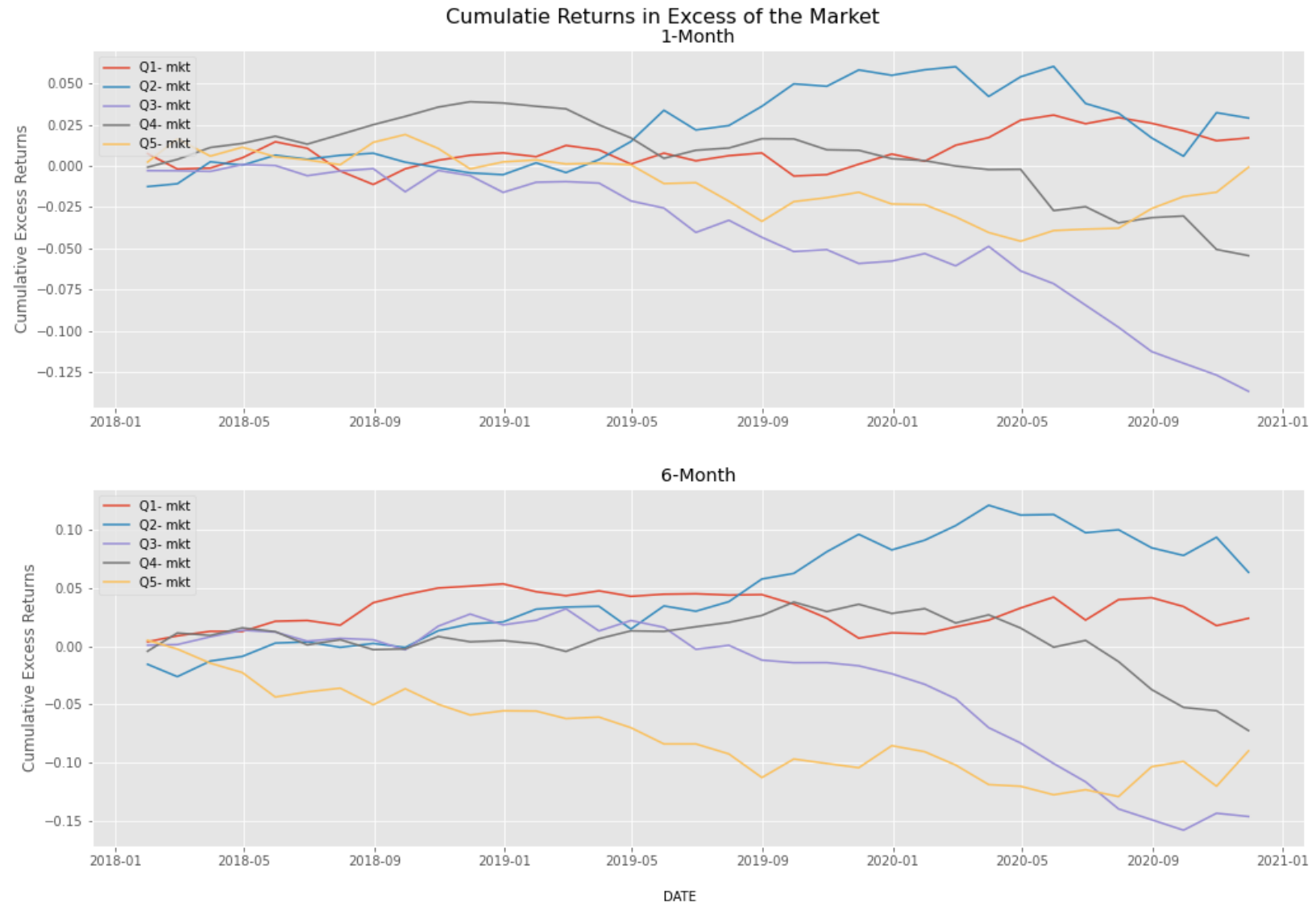
Six-Month Signal

1997-2018	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe	2019-2020	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe
Q1	11.9%	18.5%	0.54	0.64	Q1	13.2%	23.7%	0.28	0.56
Q5	7.0%	20.8%	0.25	0.34	Q5	9.2%	25.8%	0.06	0.36
Q1-Q5	4.8%	8.9%	0.33	0.54	Q1-Q5	4.0%	5.7%	0.62	0.70

Quintile Cumulative Returns 1997-2018

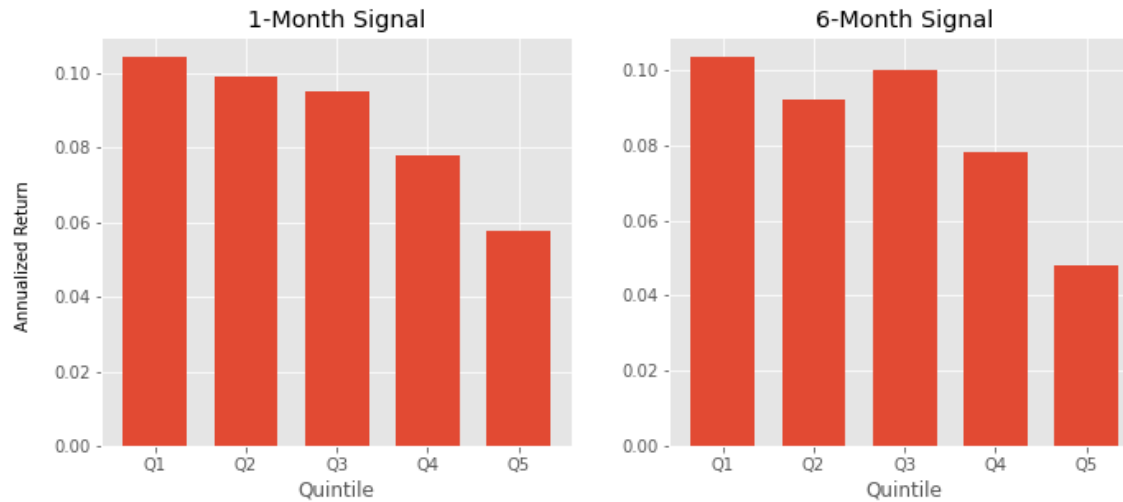


Quintiles 2019-2020 (Out-of-Sample)

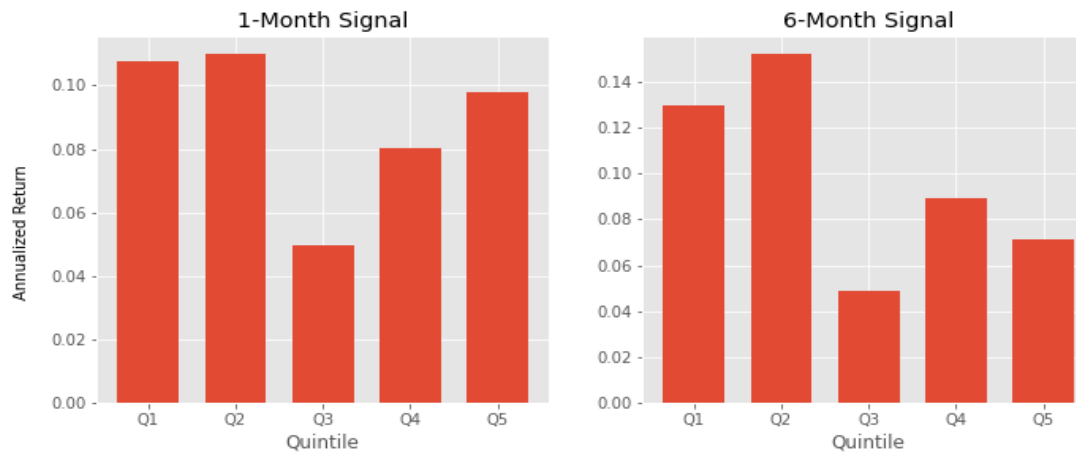


Quintile Annualized Returns

1997-2018



2019-2020



Q1-Q5 Factor Loadings (One-Month Signal)

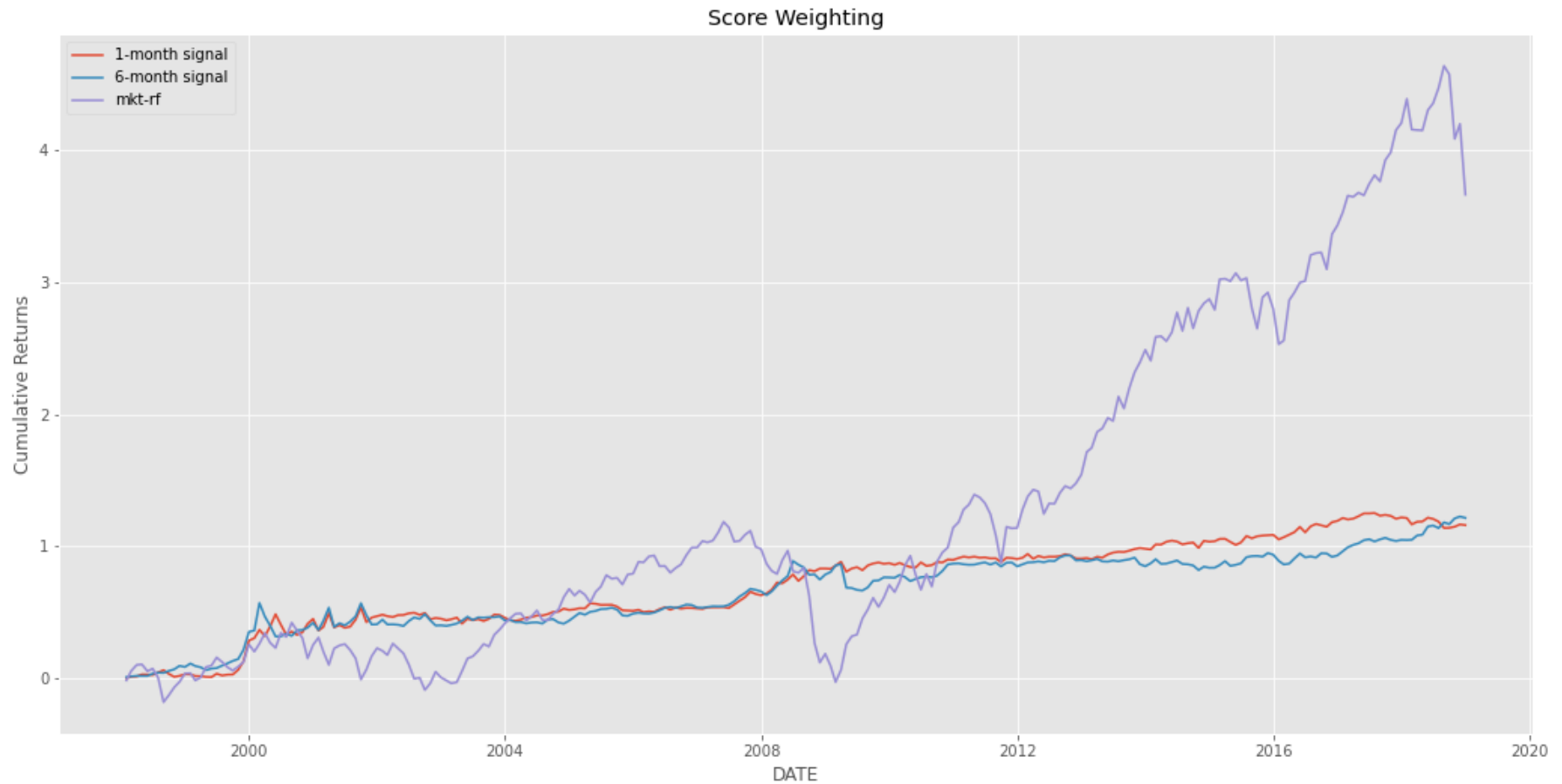
1997-2018	alpha	mkt-rf	smb	hml	mom	R ²	N
Q1	0.003	1.093	0.383	0.069	-0.039	0.917	262
	(3.209)	(44.228)	(12.000)	(2.001)	(-1.864)		
Q5	0.001	1.122	0.378	-0.060	-0.190	0.915	262
	(0.544)	(41.257)	(10.749)	(-1.580)	(-8.311)		
Q1-Q5	0.003	-0.030	0.005	0.129	0.151	0.131	262
	(1.883)	(-0.862)	(0.120)	(2.699)	(5.256)		
2019-2020	alpha	mkt-rf	smb	hml	mom	R ²	N
Q1	0.001	1.048	0.593	0.195	0.108	0.973	22
	(0.420)	(17.814)	(3.672)	(2.090)	(1.074)		
Q5	0.002	1.063	0.465	0.269	0.015	0.981	22
	(0.654)	(20.246)	(3.225)	(3.242)	(0.171)		
Q1-Q5	-0.001	-0.015	0.128	-0.075	0.093	0.239	22
	(-0.183)	(-0.291)	(0.889)	(-0.900)	(1.033)		

Q1-Q5 Factor Loadings (Six-Month Signal)

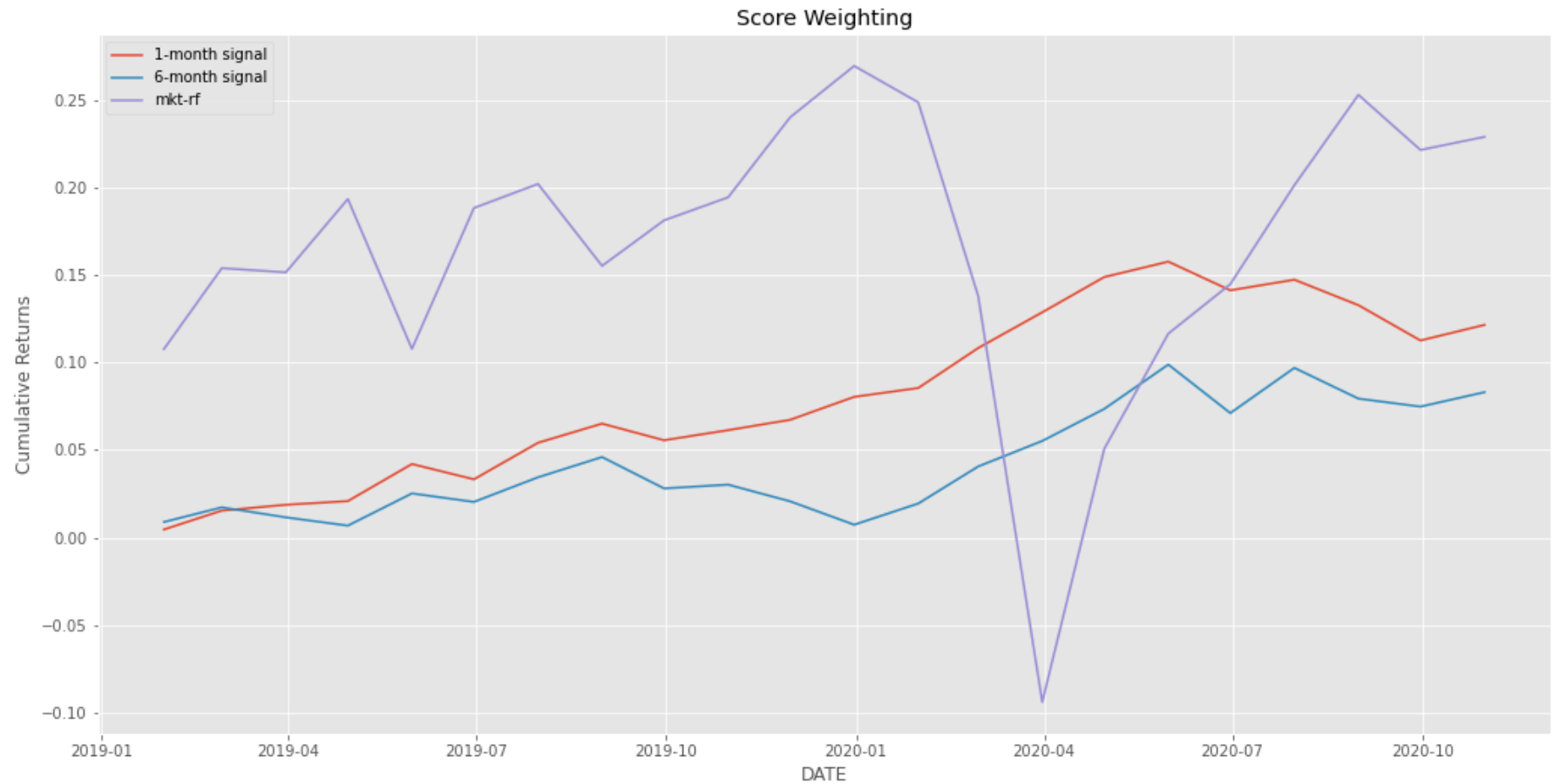
1997-2018	alpha	mkt-rf	smb	hml	mom	R ²	N
Q1	0.004	1.060	0.355	0.107	-0.017	0.930	252
	(4.268)	(47.372)	(12.244)	(3.531)	(-0.903)		
Q5	0.001	1.106	0.205	0.073	-0.279	0.911	252
	(0.794)	(38.977)	(5.581)	(1.880)	(-11.883)		
Q1-Q5	0.003	-0.047	0.150	0.035	0.263	0.365	252
	(2.248)	(-1.447)	(3.559)	(0.789)	(9.779)		

2019-2020	alpha	mkt-rf	smb	hml	mom	R ²	N
Q1	-0.001	1.107	0.373	0.218	0.128	0.968	22
	(4.268)	(47.372)	(12.244)	(3.531)	(-0.903)		
Q5	-0.001	1.115	0.271	0.290	-0.045	0.983	22
	(-0.177)	(21.605)	(1.912)	(3.550)	(-0.507)		
Q1-Q5	-0.001	-0.009	0.102	-0.072	0.173	0.326	22
	(-0.182)	(-0.129)	(0.563)	(-0.685)	(1.518)		

Score-Weight Portfolio 1997-2018



Score-Weight Portfolio 2019-2020 (Out-of-Sample)



Score-Weighting Return Statistics (Annualized)

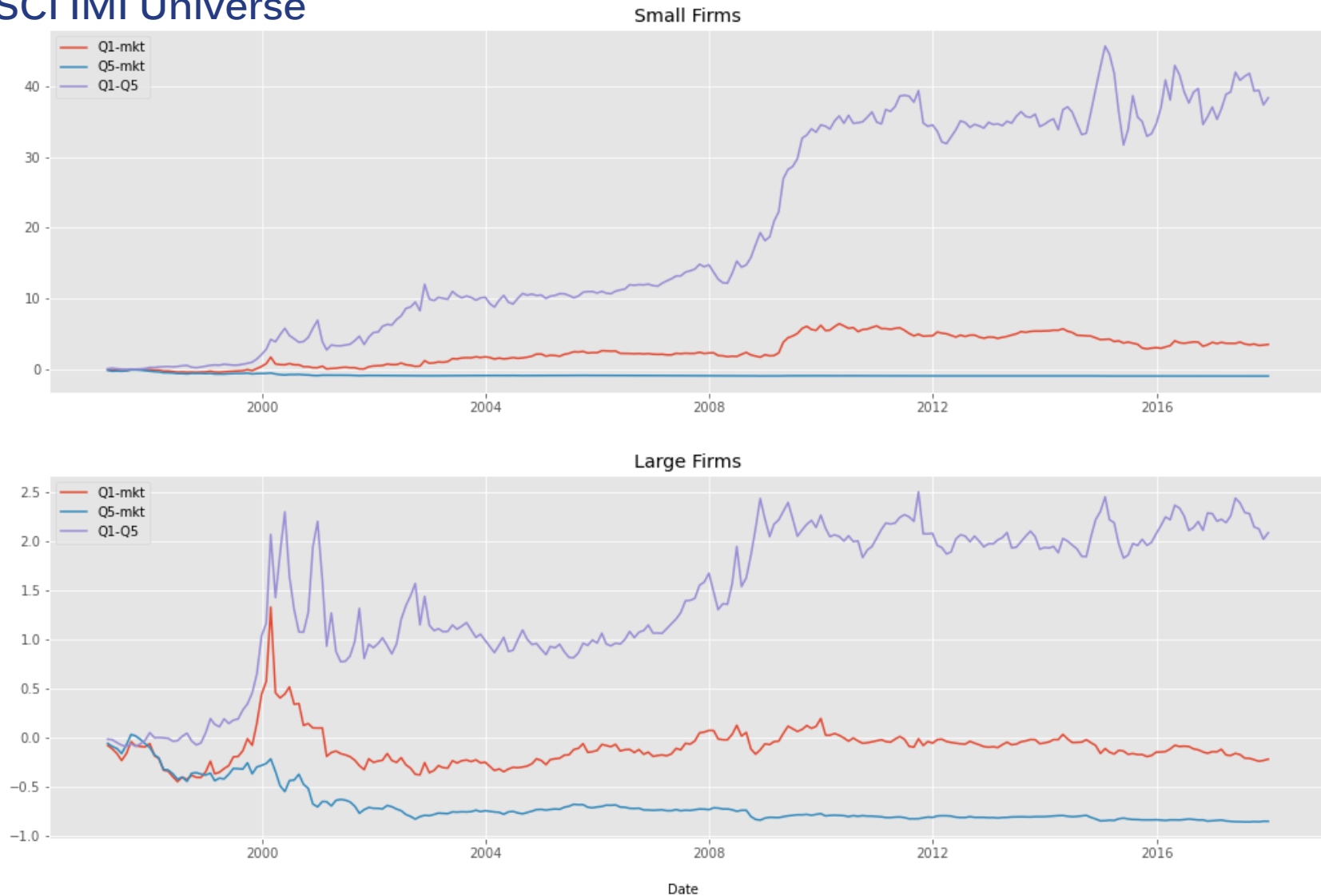
1997-2018	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe
1-Month Signal	4.0%	6.9%	0.29	0.58
6-Month Signal	4.1%	7.8%	0.28	0.53

2019-2020	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe
1-Month Signal	6.4%	4.0%	1.24	1.60
6-Month Signal	4.4%	5.0%	0.62	0.88

Robustness Tests

Effect of Capitalization (One-Month Signal)

MSCI IMI Universe



Effect of Capitalization (One-Month Signal)

MSCI IMI Universe

Small Capitalization Stocks (Returns Statistics Annualized)

1997-2018	Mean Return	SD	Median	Sharpe (excess of riskfree)	Sharpe
Q1	23.6%	31.3%	19.9%	0.69	0.75
Q5	2.0%	33.4%	7.6%	0.00	0.06
Q1-Q5	21.6%	28.1%	15.1%	0.70	0.77

Large Capitalization Stocks (Return Statistics Annualized)

1997-2018	Mean Return	SD	Median	Sharpe (excess of riskfree)	Sharpe
Q1	13.4%	20.7%	12.7%	0.55	0.65
Q5	5.2%	24.9%	11.0%	0.13	0.21
Q1-Q5	8.3%	24.1%	6.5%	0.26	0.34

Other Signal Construction Variations

- **Value-weighted** industry and fundamental returns used for signal
- **Inverse value-weighted** industry and fundamental returns used for signal
- Construct a distance metric from similarity scores and use **hierarchical clustering** to place firms into clusters--this creates a transitive fundamental group
- All the variations above are quintile sorts. We long Q1 and short Q5.
- **Remove firms of the same industry from the fundamental group** and calculate equally-weighted returns. Use these returns as the signal. For this variation long the largest values of the signal Q5 and short the smallest Q1.

Other Variations Return Statistics 1997-2018

One-Month Signal (Return Statistics Annualized)

Long-Short Position	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe
Value-Weighted	1.7%	5.0%	-0.06	0.34
Inverse Value-Weighted	2.0%	4.6%	0.00	0.43
Clustering	-6.0%	15.4%	-0.52	-0.39
Fundamental Excluding Industry	8.0%	22.1%	0.28	0.36

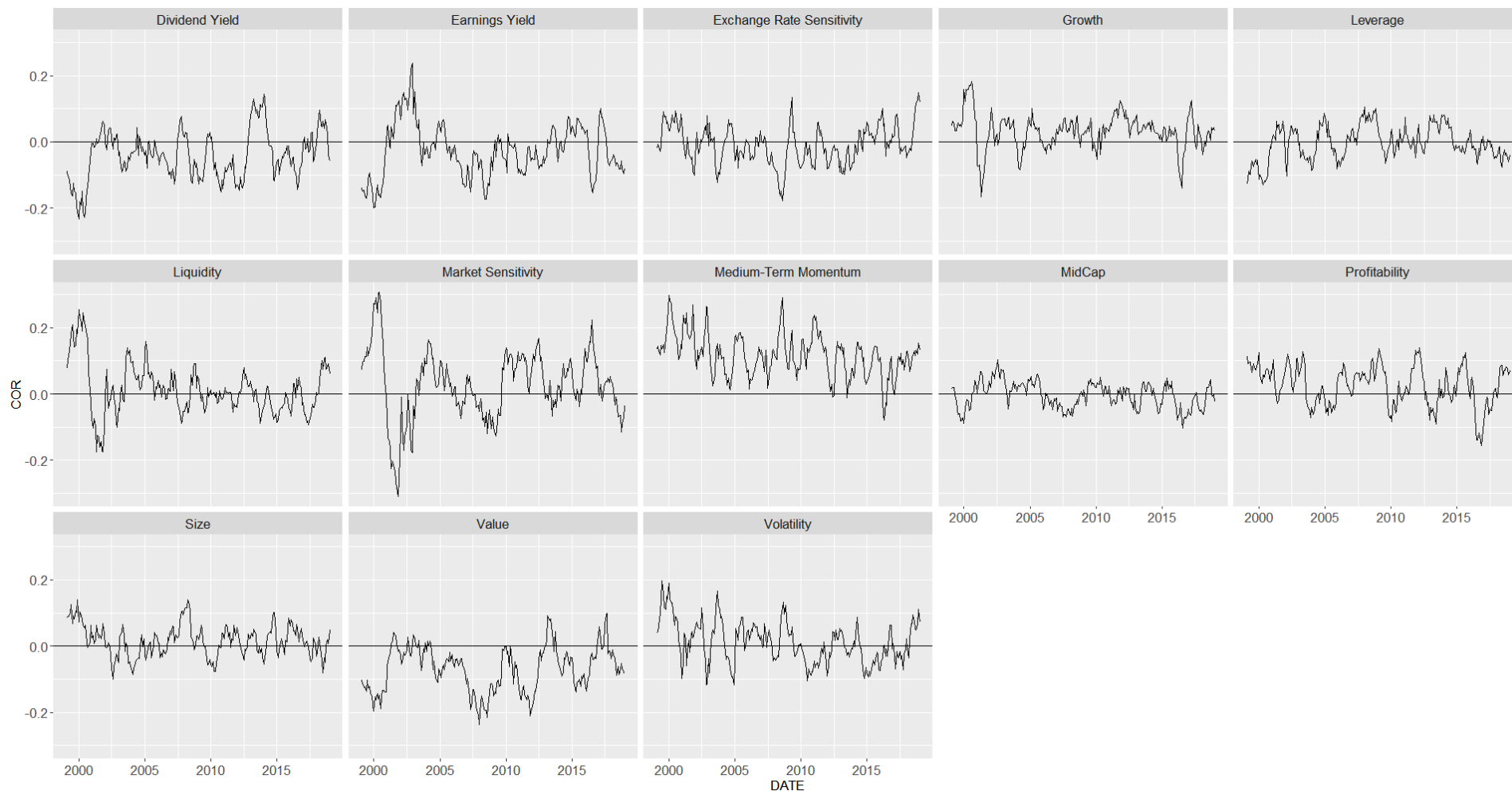
Six-Month Signal (Return Statistics Annualized)

Long-Short Position	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe
Value-Weighted	3.4%	4.9%	0.29	0.74
Inverse Value-Weighted	3.1%	4.8%	0.27	0.65
Clustering	-3.4%	15.8%	-0.33	-0.22
Fundamental Excluding Industry	6.8%	20.6%	0.25	0.33

Comments from Eric

Axioma Risk Factor Correlations

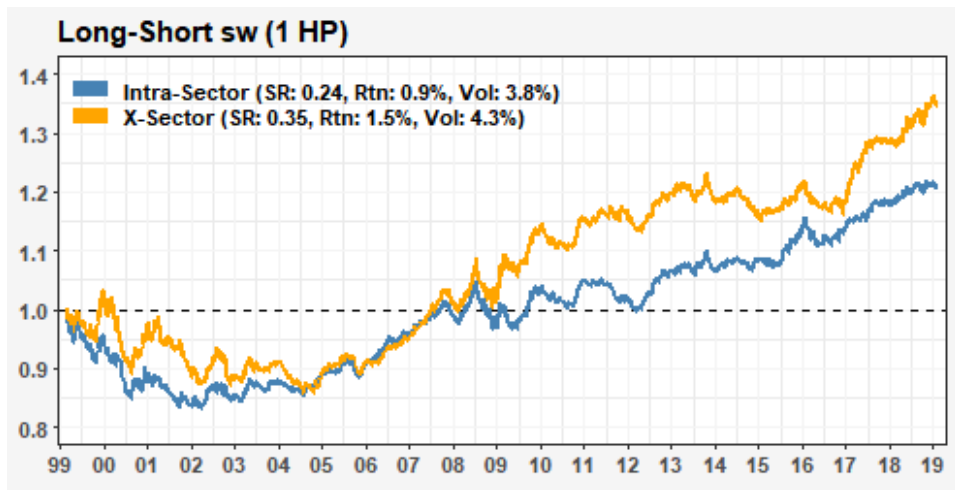
- The signal typically has positive correlations to Momentum, Market Sensitivity, Growth, and Profitability, with negative correlations to Value factors.



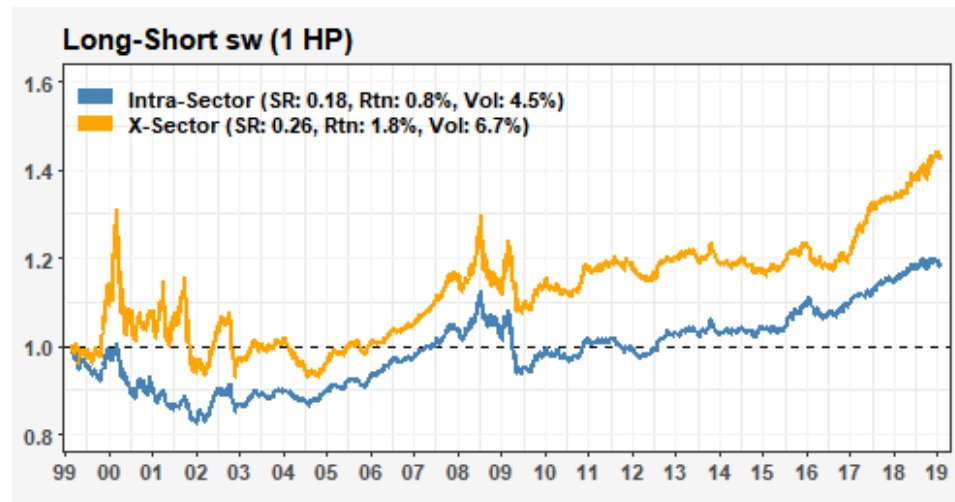
Neutralized Signal Performance

- We try neutralizing the signal's exposures to Axioma Momentum, Size, Liquidity, and Market Sensitivity through orthogonalization.
- The resulting naturalized signal shows smoother performance, especially during the tech bubble and GFC periods.
- This results in an improvement in SR to 0.35.

Neutralized Signal

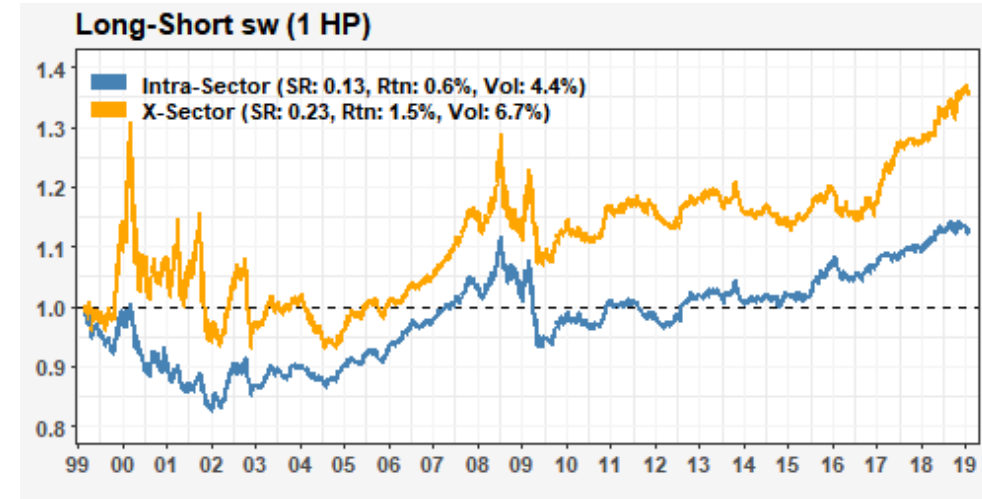
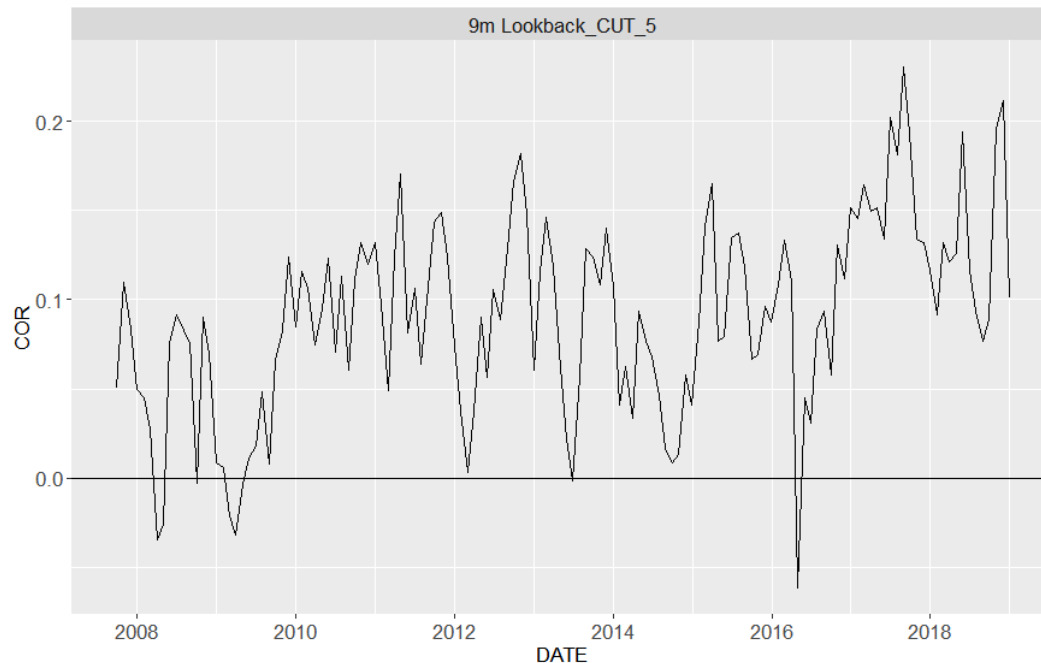


Original Signal



Correlation to Econlink Subcomposite

- The signal has on average 0.09 score correlation to SIG's Econlink factor, with some increase in recent years.
- After neutralizing the signal to Econlink, we find that the score weight performance is mostly unaffected and still shows moderate performance.
- Based on the intuition behind the signals, and these results, we would suggest incorporation into the Econlink bucket of signals.



Conclusion

Conclusion

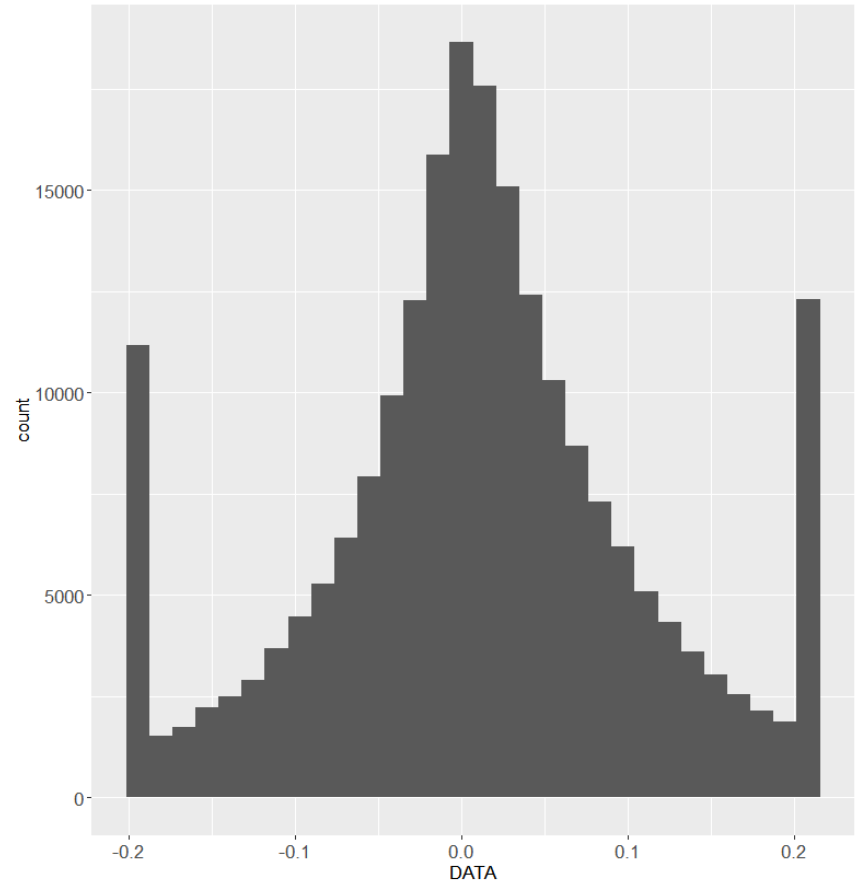
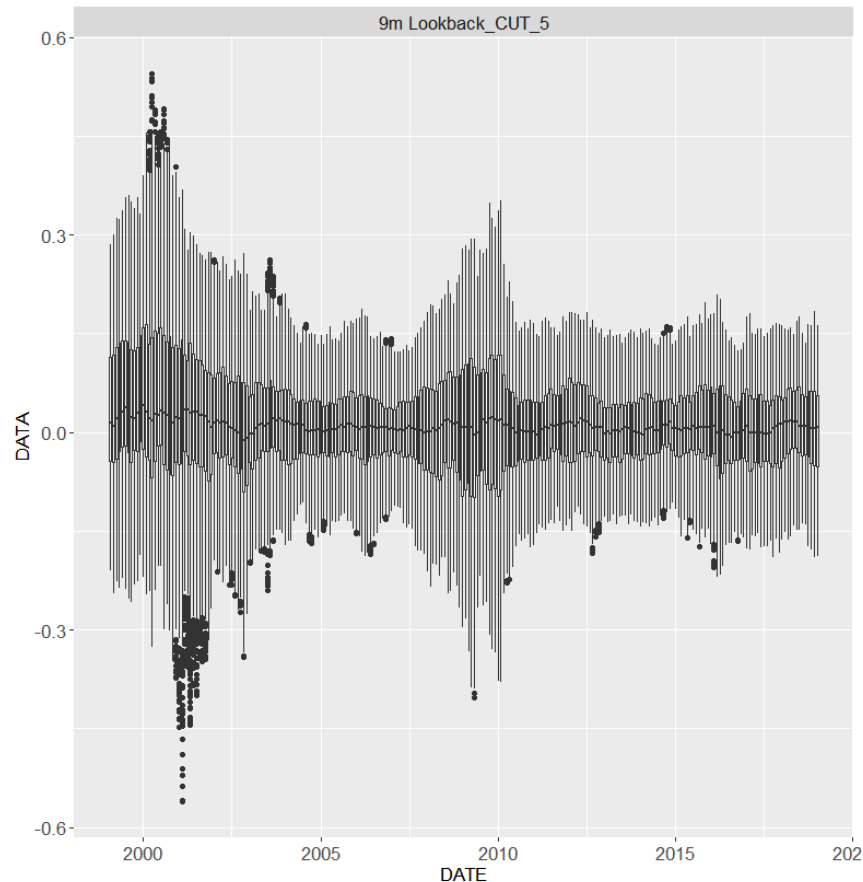
- Replication of Kruger strategy was somewhat successful—lower returns but the results make sense given difference in universe and our analysis of capitalization.
- Both in- and out-of-sample results within the SIG universe showed moderate and consistent return statistic values.
- Extending the look-back window reduces signal turnover and does not substantially affect its predictive power.
- The signal has positive exposure to Momentum and small positive correlation to SIG's EconLink subcomposite.
- Further research:
 - Build in-house 10-K similarity—try more advanced NLP techniques.
 - Signals that remove short-term firm momentum instead of industry momentum.
 - Smoothing the signal using exponential weighting or a filtration algorithm to compute longer lookback momentum.
 - Consider construction of the strategy outside the U.S.



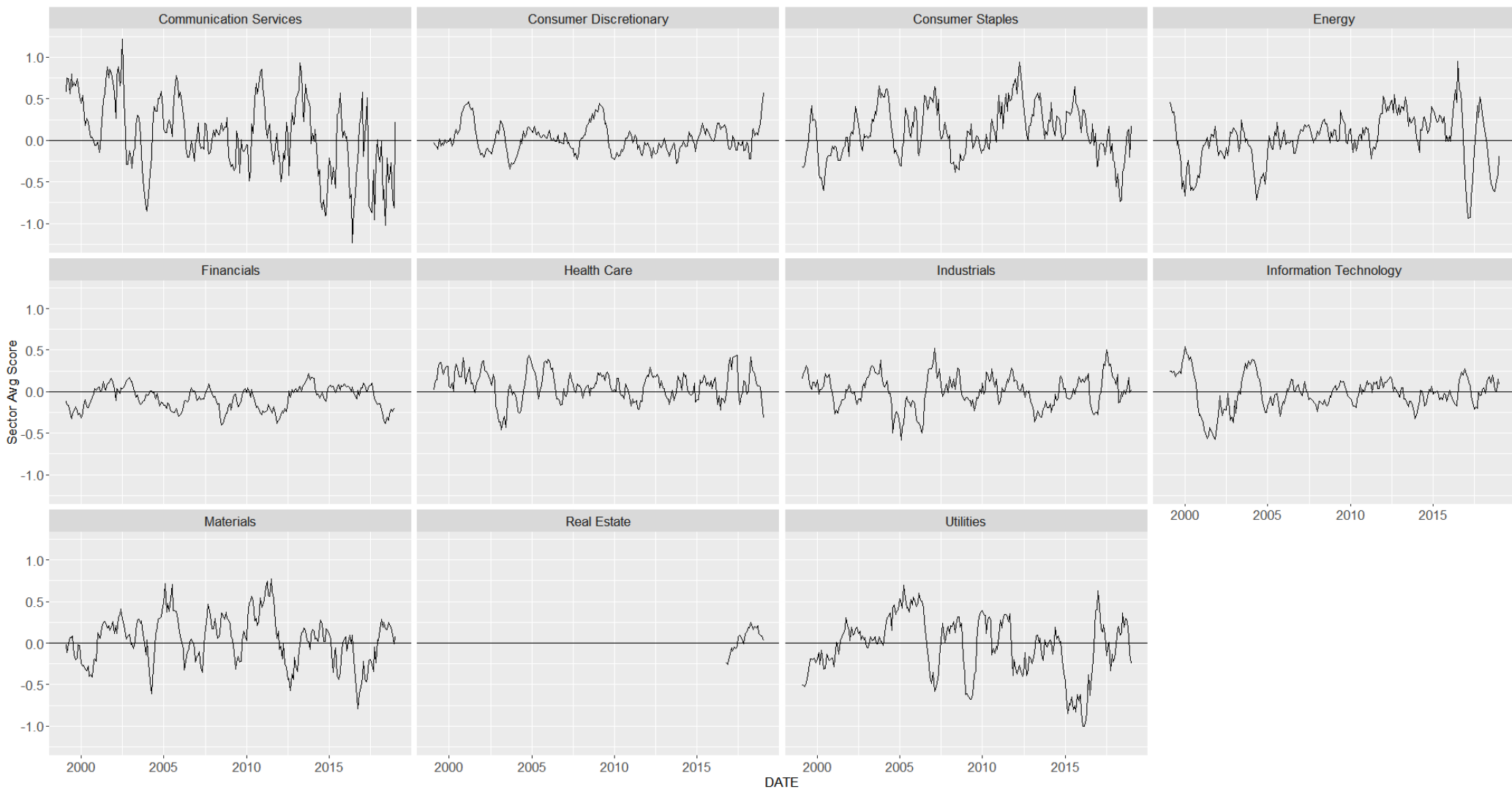
APPENDIX

Signal Distribution

- Raw signal distribution is fairly symmetrical and centered near 0, with a handful of outliers
- Increased dispersion during tech bubble and GFT periods.
- Signal will undergo rank + inverse norm transformation to form alpha



Sector Tilts



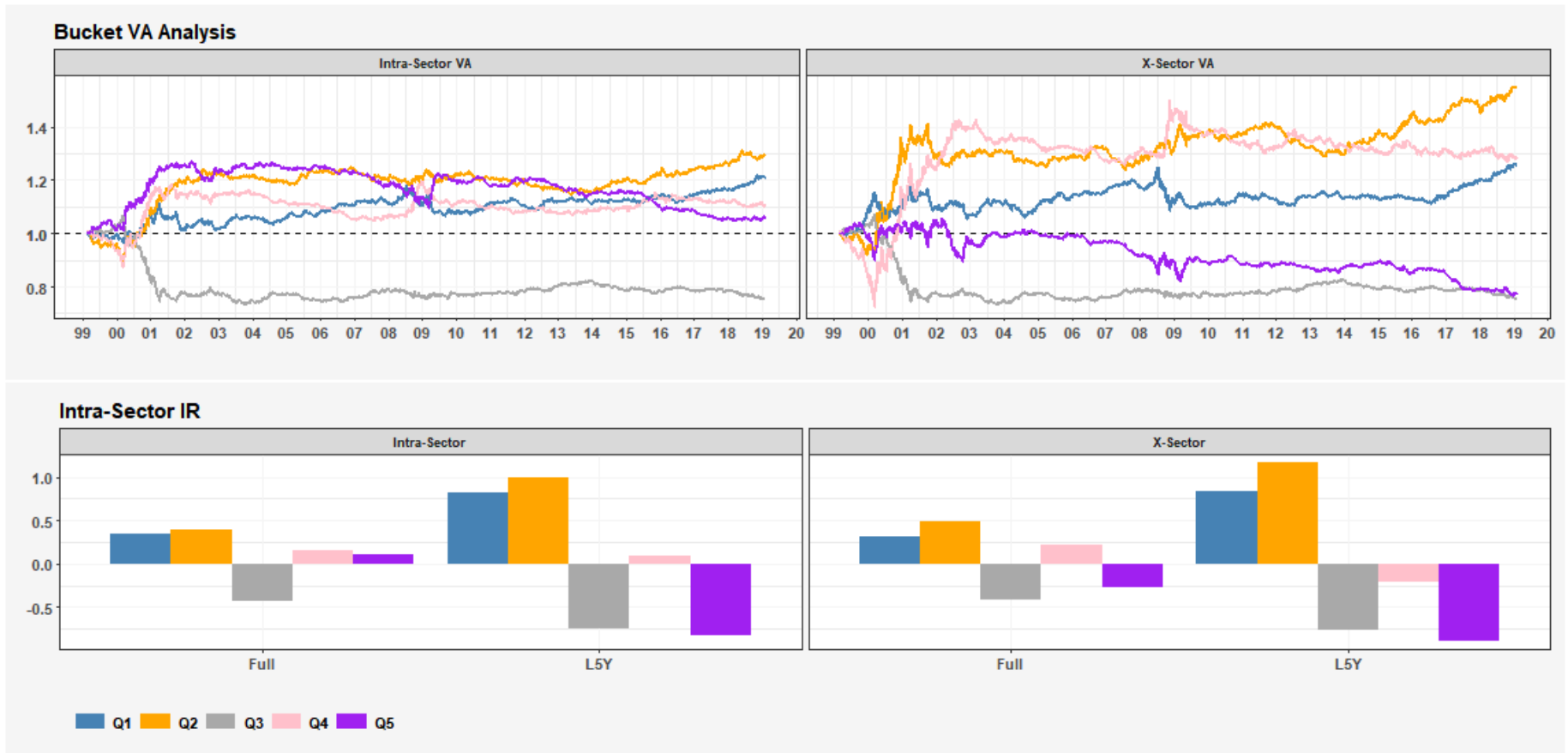
Sector ICs

- The signal's performance is weaker intrasector than fully cross sectional
- Most sector's show positive ICs except Energy, Communication Serv., and Healthcare
- Large dips in IC seen in 2000, 2009, and 2016.



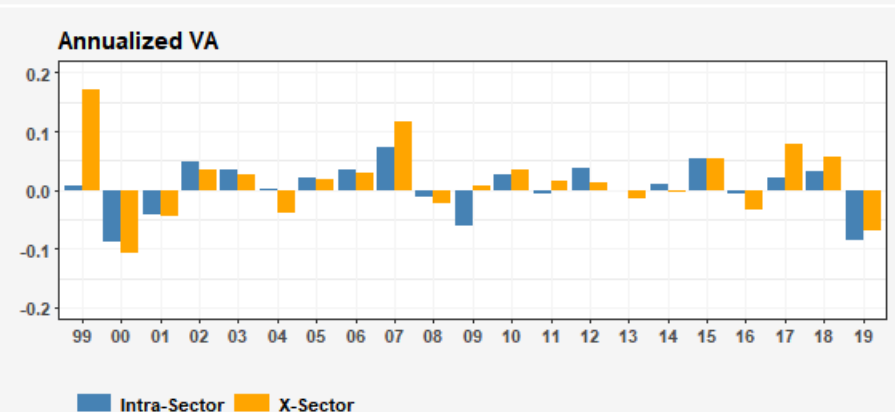
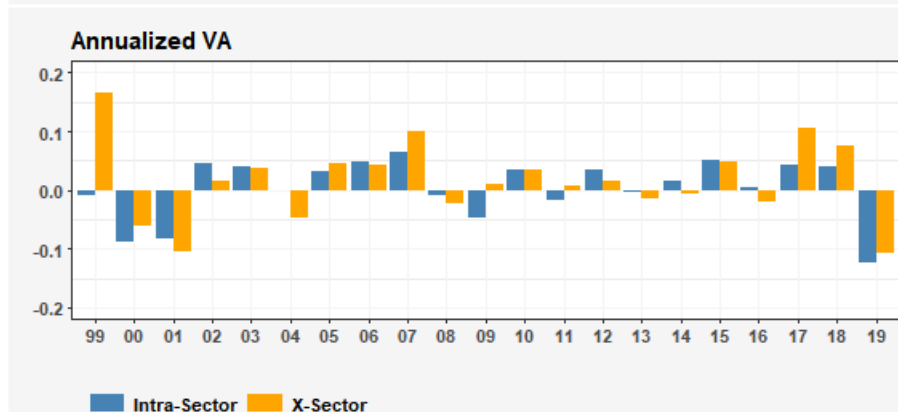
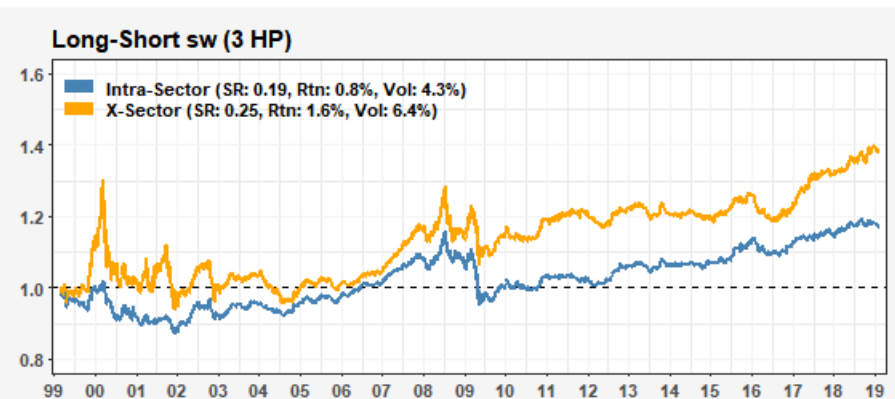
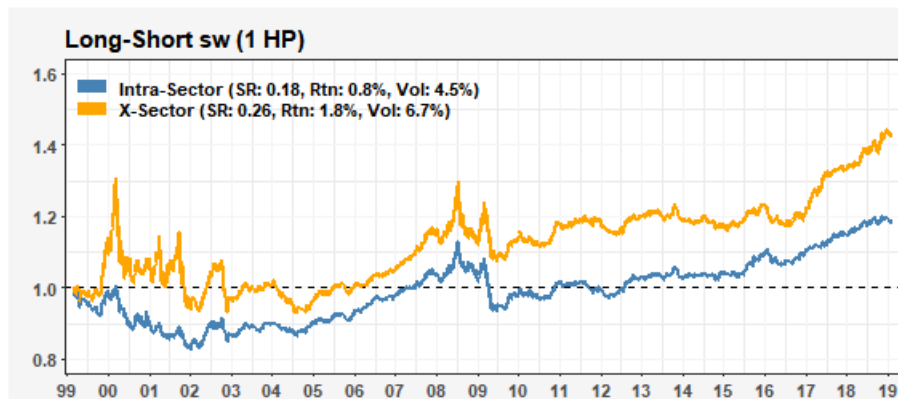
Quintile Performance

- The signal does not behave monotonically as Q2 outperforms Q1 and Q4 outperforms Q3.
- However this may largely be due to outsized volatility in the quintiles during the tech bubble period.



Score Weight Portfolio Performance

- The signal shows moderate score-weight cross-sectional performance with SR of 0.25.
- The intrasector portfolio shows lower volatility, returns, and SR.
- We see drawdowns in early 2000's, 2009, and 2016.
- With 3 month holding period, we still find similar performance for the signal.

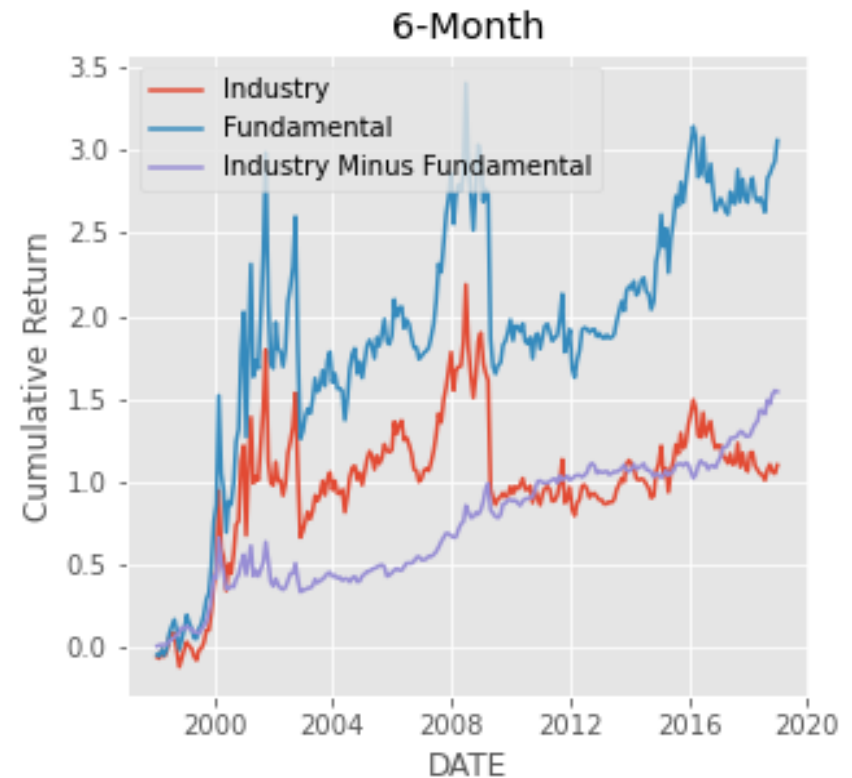
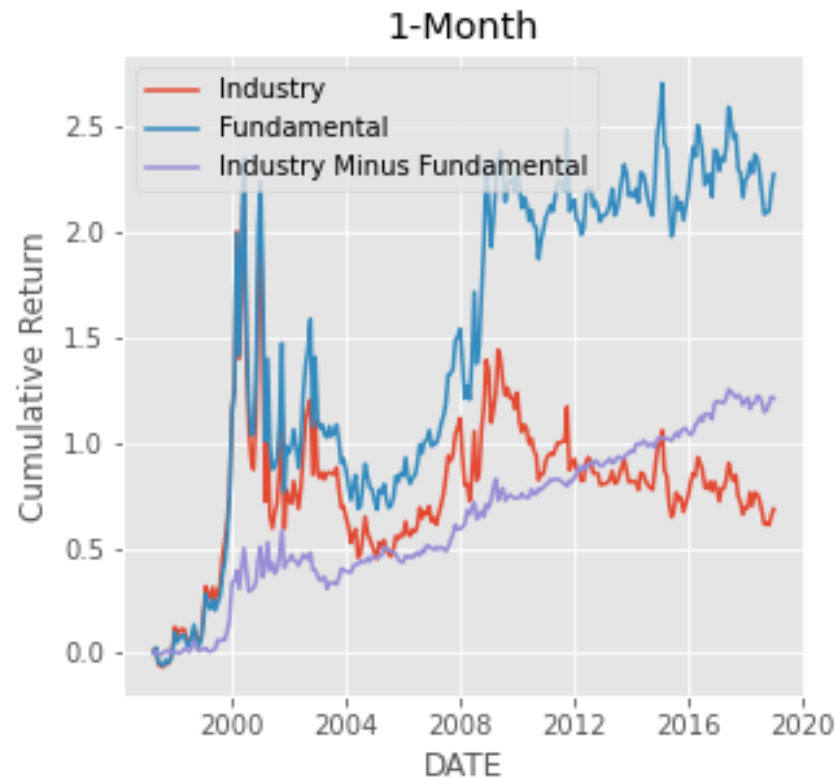


Fundamental and Industry Momentum

We considered signals based on cumulative values of $r_j^{(o)}$ and $r_j^{(f)}$. We will show results for one- and six-month signals.

- Break stocks into quintiles based on the value of the signal.
- Long Q5 — the largest values of the signal.
- Short Q1—the smallest values of the signal.

Fundamental and Industry Momentum



Return Statistics 1997-2018

One-Month Signal (Return Statistics Annualized)

	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe
Industry	5.0%	23.0%	0.13	0.22
Fundamental	8.3%	23.7%	0.26	0.35

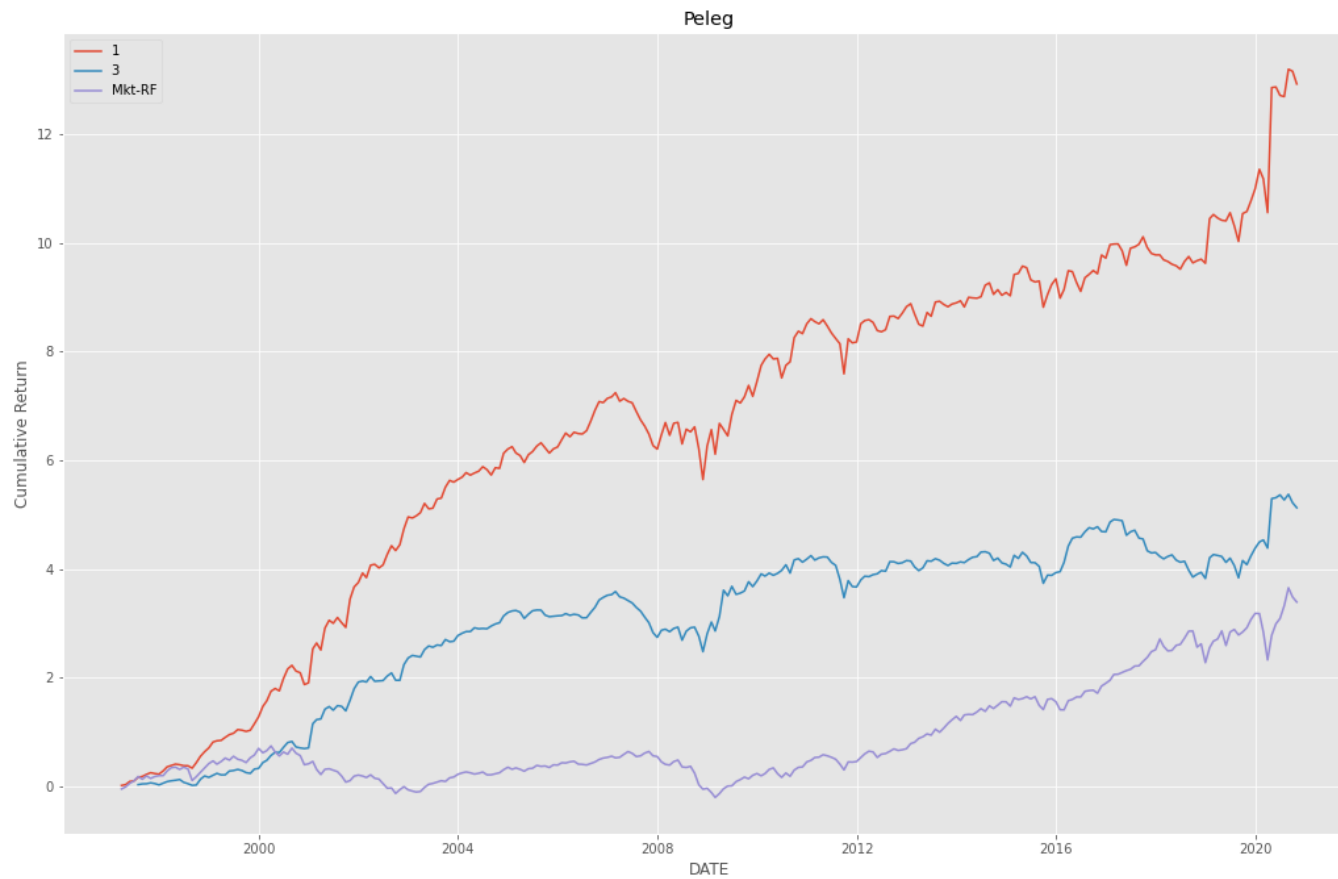
Six-Month Signal (Return Statistics Annualized)

	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe
Industry	6.0%	22.1%	0.19	0.27
Fundamental	9.4%	23.1%	0.32	0.41

Peleg Strategy

- The Peleg strategy is a variation of short-term reversal.
- It invests in stocks based on cumulative values of $r_j^{(f)} - r_j$.
- Break the signal into quintiles based on the difference.
- Long the quintiles corresponding to the largest values Q5.
- Short the quintiles corresponding to the smallest values Q1.

Peleg Strategy 1997-2018



Peleg Return Statistics 1997-2018

	Mean Return	SD	Sharpe (excess of riskfree)	Sharpe
1-Month Signal	11.8%	11.3%	0.87	1.04
3-Month Signal	8.4%	11.7%	0.57	0.72

