Machine Learning in Finance Workshop 2021

Why and how systematic strategies decay

Adam Rej

Capital Fund Management International Inc.

Columbia University Data Science Institute

Columbia Engineering | The Fu Foundation School of Engineering and Applied Science

Bloomberg

- A lot of time and energy has been devoted to disproving the Efficient Market Hypothesis.

- A common (but of course not unique) approach is to propose strategies that yield statistically significant excess returns in the backtest and with the expectation that they will continue to do so in the future.

- Many such strategies have been proposed. Hou et al. (2018) documents 400+… a zoo of strategies

- Are these genuine findings or data mining? Do market participants arbitrage them away?

- Several authors (Harvey et al. (2015), McLean and Pontiff (2016), Chen and Zimmerman (2020)) have attempted to quantify what happens to those strategies after publication.

- They find a substantial degradation of out-of-sample performance which they attribute to:

  - overfitting (data mining)
  - arbitrage capital

- In this talk I would like to discuss how to detect / differentiate each effect

- Talk based on a working paper together with Antoine Falck and David Thesmar, "*Why and how systematic strategies decay*"

- We have recreated (recoded) 72 equity strategies published in the financial and accounting literature. Many of those are familiar to you: **size, book-to-market, momentum**, etc.

- Having the code + portfolio positions at our disposal, we are able to propose (cross-sectional) variables that proxy for overfitting and arbitrage.

- We find that several of them are statistically significant.

Our zoo of strategies

- We replicate 72 long-short strategies (a.k.a factors).

- All strategies were ***originally*** proposed using US stock data only (CRSP and Compustat).

- We only took strategies published before 2011 in order to have sufficient OOS period.

- We follow the original strategy recipes "line by line". However we "debias" the final PnL using a 36m rolling window. This ensures the absence of residual market exposure

- Thanks to CFM's proprietary data sets, we are able to seamlessly define these strategies on international pools: **Australia, China, Europe, Hong Kong, Japan, South Korea, UK**

- We remove strategies whose in-sample Sharpe is < 0.3

	Entire sample		**In-sample Sharpe > 0**		**In-sample Sharpe > 0.3**	
	Sharpe ratio	*t*-stat	Sharpe ratio	*t*-stat	Sharpe ratio	*t*-stat
Mean	0.98	4.52	1.02	4.73	1.15	5.34
Median	0.85	3.79	0.88	3.84	0.99	4.55
Q_1	0.43	1.89	0.46	1.98	0.69	2.75
Q_3	1.33	6.24	1.38	6.39	1.46	7.10
N	72		69		60	

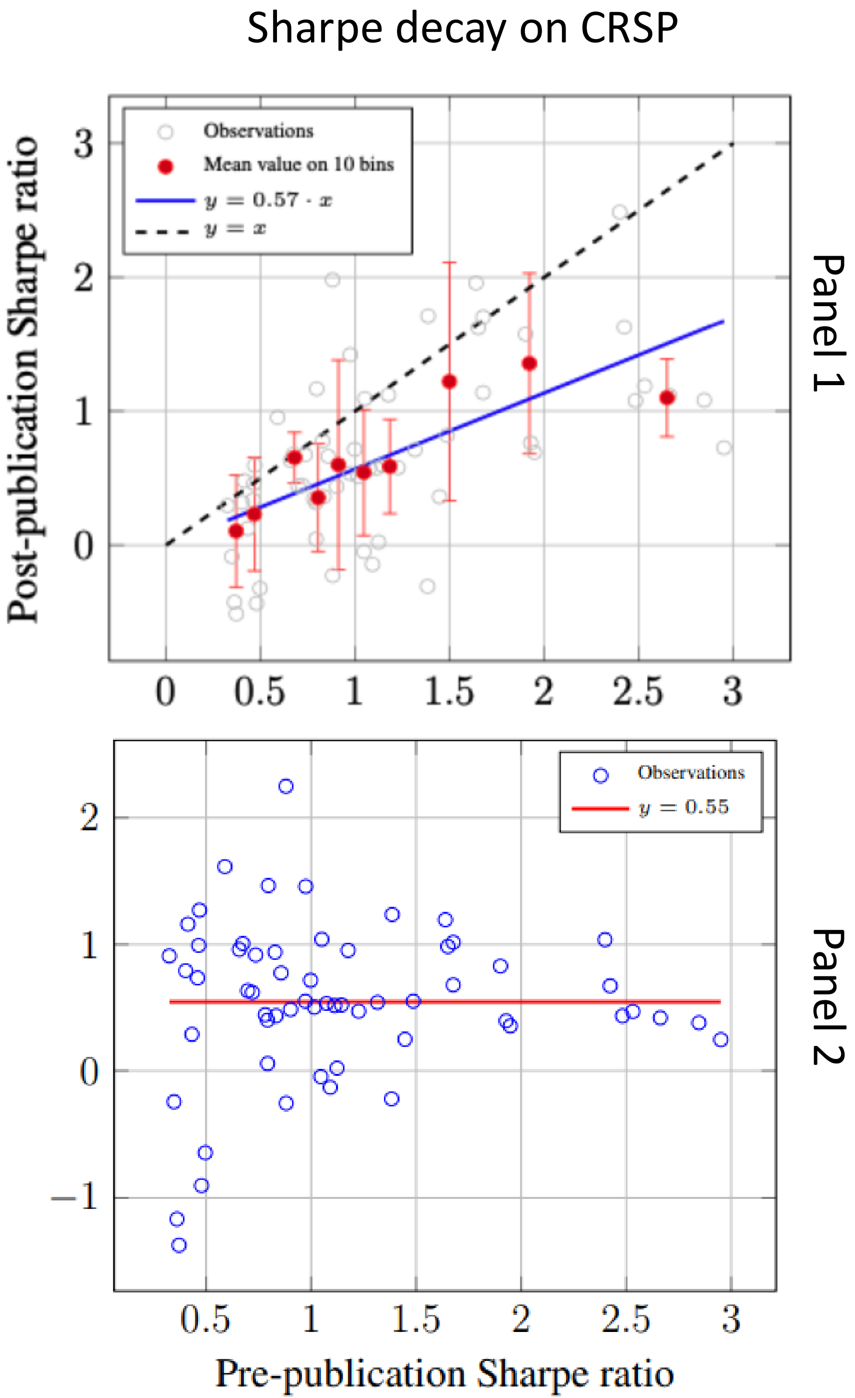
Out-of-sample decay: part 1

- We capture the decay using the discount factor

$$q = \frac{SR_{OOS}}{SR_{IS}}$$

- Bouchaud, Rej and Seager (2019) have proposed a simple model of overfitting tailored to investment research. For strategies discussed in this talk, the model predicts q of just above 0.5

- This is very much in line with what happens to factors in our zoo and in line with McLean and Pontiff (2016), who find a slightly bigger drop on a different set of factors



Out-of-sample decay: part 2

- There is yet another way of out-of-sample testing of these strategies…

- … by evaluating their performance on international pools. Data not looked at by the authors!

- So the entire performance on these pools is out-of-sample. Wait, but what’s the in-sample performance? CRSP!

$$q =? \frac{SR_{pool}}{SR_{IS}^{CRSP}}$$

- This new definition of q is unfortunately flawed, because international pools tend to be much smaller than CRSP (few thousand stocks) and this will bias q to be artificially low.

- We need to adjust for the size of the pool.

- Let's assume a simple model of returns in the form

$$r_{t+1} = (b + \eta_{t+1}) s_t + \beta R^m_{t+1} + \epsilon_{t+1} \qquad (1)$$

- This is essentially an extension of CAPM, where s_t is a predictive characteristic and b, β are scalar loadings. η_t on the other hand captures shocks.

- The Sharpe ratio of the strategy is given by

$$SR = \frac{b}{\sigma_\eta} \frac{1}{\sqrt{1 + \frac{12\sigma_\epsilon^2}{\sigma_\eta^2 N}}} \qquad (2)$$

- In the absence of eta shocks, this implies square-root dependence on the pool size N. With shocks however the dependence is

$$SR \approx \underbrace{\frac{b}{\sigma_\eta}}_{SR_\infty} \left(1 - \frac{6\sigma_\epsilon^2}{\sigma_\eta^2} \frac{1}{N}\right) \qquad (3)$$

- We will use both ways of adjusting for size (without and with the eta shock). The latter requires bootstrap to determine the coefficients

Country	**Index/Pool**	**Average size**	**Raw**	**Size adjusted**	
				Simple	**Complex**
United-States	CRSP post-publication	4694	0.55	0.52	0.54
Australia	S&P/ASX 200 Index	200	0.08	0.38	1.00
Europe	Bloomberg European 500	505	0.09	0.24	0.37
Hong Kong	Hang Seng Composite Index	312	0.20	0.75	0.74
Korea	Korea Kospi Index	734	0.20	0.52	0.44
China	China SE Shang Composite	918	0.18	0.39	0.32
Canada	S&P/TSX Composite Index	244	0.11	0.47	0.87
Japan	TOPIX 500 Index (TSE)	500	−0.01	−0.02	0.32
United Kingdom	FTSE 100 Index	100	0.08	0.53	1.78
Average non-US		439	0.11	0.41	0.73

The average of the two gives 0.57

Determinants of decay

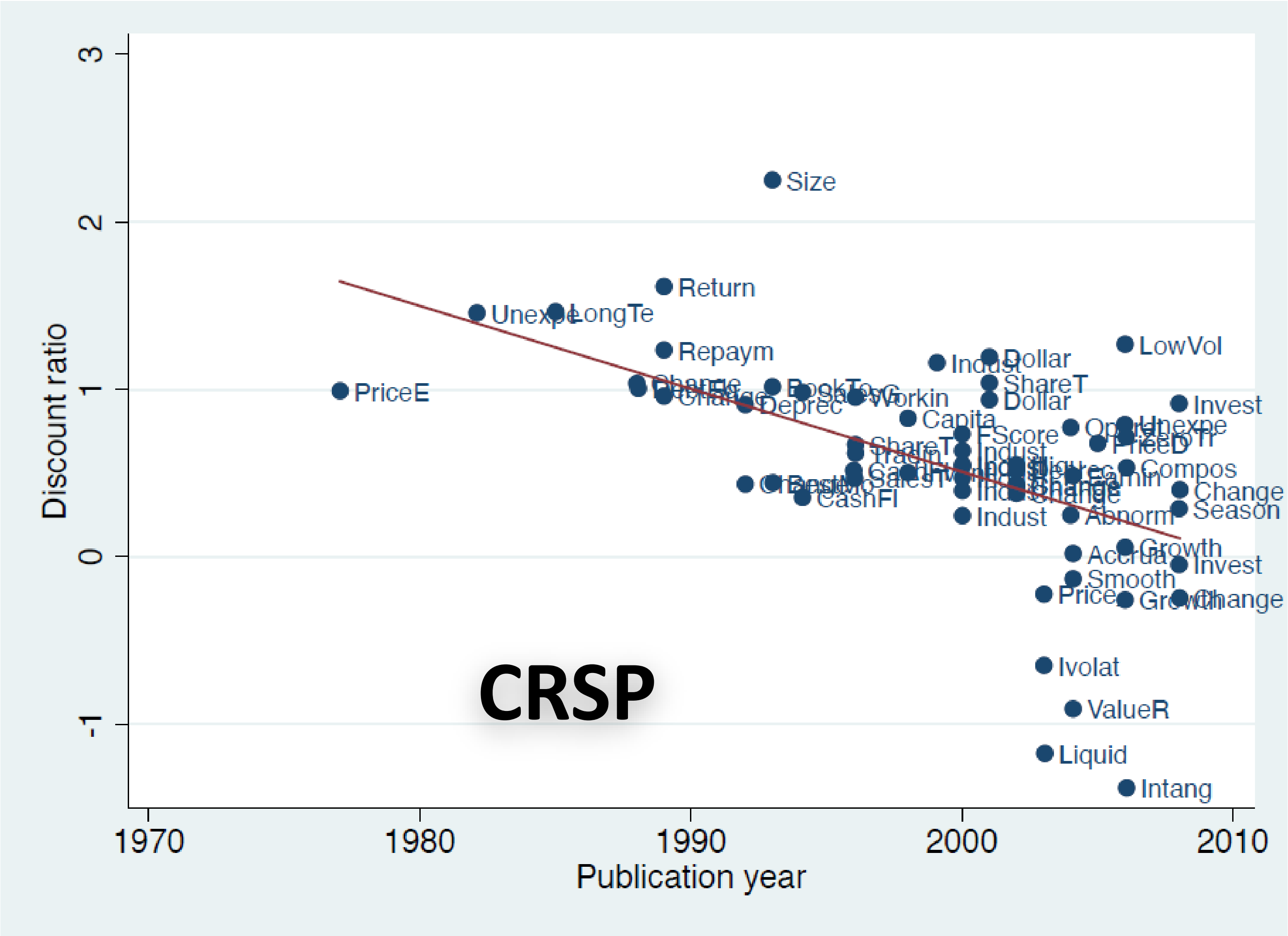
- OOS decay as function of publication date.

- Coherent with both hypotheses behind the decay:

  More arbitrage capital rush in recently (arbitrage)

  Low-hanging fruits have been plucked, so people try harder and harder (overfitting)

- Can we determine factors driving the decay?



Determinants of decay: arbitrage view

- We propose proxy variables related to broadly-defined liquidity of the stocks used in the portfolio. We work with CRSP.

- The more illiquid stocks in the portfolio, the more difficult it should be to arbitrage away the strategy:

1. Amihud's liquidity

$$Amihud_{s,t} = -\frac{|r_{s,t}|}{V_{s,t}}$$

stock return

volume traded

   computed per stock each day then averaged and weighted by absolute stock weights in the portfolio

2. Log market cap of factor (the smaller market cap, the more difficult for arbitrage capital to move in in size)

3. Log market cap of the short leg (should capture the difficulty in shorting stocks)

4. Holding period aka turnover

5. BA?

- We perform univariate cross-sectional regressions
- I remind you that we have discarded strategies with SR<0.3
- All regression signs are negative.
- 2/4 variables are statistically significant.

CRSP

const	0.87	0.70^{***}	0.72^{***}	0.36^{***}
	(1.22)	(7.8)	(7.86)	(2.64)
log holding period	−0.05			
	(−0.41)			
log mkt cap long short		-0.14^{**}		
		(−2.48)		
log mkt cap short			-0.15^{***}	
			(−2.67)	
liquidity				-0.34^{*}
				(−1.95)
R^2	0.00	0.10	0.11	0.06

$^{***}\ p < 0.01$; $^{**}\ p < 0.05$; $^{*}\ p < 0.1$

Determinants of decay: overfitting view

- We group the variables pertaining to the overfitting view into 2 groups:

Researcher incentives

1. t-stat < 3 (1)
    - t > 3 reduces multiple testing problems (Harvey et al., 2015)
    - Bonferroni; 5 independent hyp tests with t>3 equivalent to 1 with t>2.5

2. Quantile flexibility (2)
    - Long-short portfolios use author-specified quantiles, but we can use it as a param (5%, 10%, …)
    - compute $\frac{\mathrm{std}(SR_q)}{SR_{in}}$
    - variant : $\max(SR_q) - SR_{in}$

3. Formula flexibility (2)
    - # of Compustat fields > 2
    - # of operations > 2

4. Publication date (people try harder)

Small sample issues

1. Short in-sample period (MINUS # of months in-sample) (1)

2. Sensitivity to random 10% of sample (1):
    - drop randomly 10% of stocks every period and compute SR_{mod}
    - repeat 100 times
    - compute $\log(\mathrm{std}(\mathrm{SR}_{mod}))$

3. Sensitivity to 0.1% of most influential observations (1):
    - every day drop 0.1% observations for which $|w_{i,t}r_{i,t}|$ is largest and re-compute the Sharpe ratio
    - compute the absolute Sharpe loss

$$|SR_{dropped\ 0.1\%} - SR_{in}|$$

- We perform univariate cross-sectional regressions
- All regression signs save for one are negative.
- 4/9 variables are significant

CRSP

	1	2	3	4	5	6	7	8	9
const	0.59***	0.52***	0.54***	0.63***	0.92***	0.55***	0.59***	0.59***	0.53***
	(6.12)	(6.09)	(6.39)	(6.29)	(5.9)	(6.73)	(8.01)	(7.99)	(7.72)
tstat<3	−0.15								
	(−0.8)								
quantile flexilbity I		−0.10							
		(−1.08)							
quantile flexibility II			−0.07						
			(−0.83)						
formula complexity (# fields)				−0.23					
				(−1.37)					
formula complexity (# operations)					−0.49***				
					(−2.72)				
- sqrt nb months is						0.08			
						(0.91)			
Sensitivity to dropping 10%							−0.16**		
							(−2.04)		
Sensitivity to influential 0.1%								−0.21***	
								(−2.99)	
Publication date									−0.35***
									(−5.06)
N	60	58	58	60	60	60	57	58	60
R^2	0.01	0.02	0.01	0.03	0.11	0.01	0.07	0.14	0.31

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Arbitrage vs. overfitting

- We aggregate overfitting variables (bar for DAPUB) into one super variable…
- …same for arbitrage proxies…
- …and we run a set of regressions
- We see that while arbitrage variables are statistically significant, they add little in terms of R2

Dependent variable:	Discount Ratio				
	(1)	(2)	(3)	(4)	(5)
Year of publication -1990	$-.05^{***}$			$-.041^{***}$	$-.044^{***}$
	(-4.9)			(-4.9)	(-5.2)
Arbitrage vulnerability		$-.28^{**}$		$-.13$	
		(-2.6)		(-1.5)	
Overfitting vulnerability			$-.34^{***}$	$-.28^{***}$	$-.31^{***}$
			(-3.1)	(-3.0)	(-3.5)
Constant	1.0^{***}	$.58^{***}$	$.55^{***}$	$.92^{***}$	$.94^{***}$
	(8.7)	(7.7)	(7.4)	(9.6)	(9.8)
N	60	58	55	55	55
R^2	0.30	0.11	0.15	0.47	0.45

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Outlook and conclusions

- We have recoded a sizeable set of strategies (72).
- This allows us to study their out-of-sample performance. We have reproduced their performance decay on CRSP in line with what other authors have found.
- Using CFM's proprietary international stock data, we define these strategies on international pools and study their decay. The original papers proposing these investment strategies did not (at least no mention of this) look at international data, so this data is out-of-sample. After accounting for different pool sizes, we find performance decay in line with that on CRSP.
- The control over the code and outputs allows us to define quantities that may capture overfitting or arbitrage effects.
- We have proposed sets of proxy variables for each of these views. Of course our sets are not exhaustive. Bid/ask spreads, for example, would be an even better liq proxy.
- We run univariate regressions to identify variables predictive of out-of-sample decay. We find statistically significant coefficients for variables from both sets.

- Market capitalization (or size-related variables) for the arbitrage set and sensitivity to the pool and big movers together with formula complexity for the overfitting view.
- Date of publication is a strong driver of out-of-sample decay, but it is unclear whether it’s because of overfitting or arbitrage.
- It would be interesting to develop more systematic tools to monitor arbitrage and overfitting. We believe we have made our modest contribution here, but probably a lot more may be done.
- You can incorporate these ideas to have a better idea of expected future returns of your own strategies!

THE END