Anomalies and Multiple Hypothesis Testing: Evidence from Two Million Strategies

Tarun Chordia

Amit Goyal

Alessio Saretto

May 2018

A zoo of factors/anomalies

Many factors

- Fama and French (2015) five-factor/ Hou, Xue, and Zhang (2015) four-factor/ Stambaugh and Yuan (2016) four-factor
  - Plus the Carhart (1997) momentum factor

Many anomalies

- Hou, Xue, and Zhang (2015) – 80
- McLean and Pontiff (2016) - 97
- Linnainmaa and Roberts (2016) – 38
- Harvey, Liu, and Zhu (2016) - 316
- Hou, Xue, and Zhang (2017) – 447

Are these reliable?

McLean and Pontiff (2016) look at performance **AFTER** publication

- Maybe academics make anomalies go away
- Chordia, Subra, Tong (2014) – hedge fund AUM

Linnainmaa and Roberts (2016) look at performance **BEFORE** time-period studied in publication

- Maybe anomalies did not exist before the time they were studied

Hou, Xue, and Zhang (2017) look at performance **DURING** the time-period

- Maybe there are no anomalies to begin with

P-hacking

Harvey, Liu, Zhu (2016) – **"… most claimed research findings in financial economics are likely false**."

- Recommend a t-statistic hurdle of 3 - **MHT**

Harvey (2017) – AFA Presidential address

- Usually, only significant results have a path to publication
- Try many variables until the significant one is found
- Try different sample periods, Different data choices
- Different test procedures
- Focus on microcap firms

Lo and MacKinlay (1990) – Data Snooping

What we do

Construct a laboratory experiment to analyze many trading strategies including:

- Strategies that have been studied and published
- Strategies that have been studied and not published (because not significant)
- Strategies that have not been studied (maybe because we cannot describe an intuitive story behind why they would work or not work)

What we do

Carefully construct a multiple hypothesis test

- Extend methods described by Harvey, Liu, and Zhu (2016)
- Use a stationary bootstrap that allows control for cross-correlation in strategy

Check which strategies survive and which do not

- Take into account statistical thresholds and impose economic thresholds

What we do

We conduct a deliberate large-scale data mining exercise

Not interested in promoting any particular strategy

- Not a fishing expedition but data mining in Leamer (1978) sense

No a priori idea of what will work or not

- Not a data snooping exercise

One can also think of our exercise as a UBER test of market efficiency

Generating strategies

Construct trading signals based on COMPUSTAT and CRSP data

- Select all items of COMPUSTAT that are sufficiently populated (at least 300 firms each year for at least 30 years)
- Trading signals based on CRSP: price, size, total volatility, previous returns (1 to 12 months), turnover, volume
  - Obtain 168 variables
- construct signals: levels and growth rates, ratios of two levels or growth rates, x_1/x_2, and all permutations of three variables, $(x_1 - x_2)/x_3$
- Total of 2,385,778 signals

Generating strategies

Require very stringent filters to avoid generating strategies that would not be tradable

- At least 6 months between portfolio formation (June 30) and timestamp on COMPUSTAT data
- Remove all stocks with price less than $3 on June 30
- Remove all stocks in the bottom quintile of NYSE market cap distribution

Sample period is 1972 to 2015

- Increase in number of firms due to Nasdaq

Strategy evaluation (alpha and lambda)

10−1 decile portfolio alphas from FF5+MOM

$$R_{pt} = \boldsymbol{\alpha} + \beta' F_t + e_{pt}$$

FM coefficients

$$R_{it} - \widehat{\beta}_i F_t = \lambda_{0t} + \boldsymbol{\lambda}_{1t} X_{it-1} + \lambda_{2t} Z_{it-1} + u_{it}$$

- Risk-adjusted returns on LHS (Brennan, Chordia, and Subrahmanyam, 1998)
- Z's are control variables (R1, R212, Sz, B/M, Profitability, Investment)

Summary statistics

	Average return									
	N	Mean	Median	Std	Min	Max	|ret| > 0.5%		|ret| > 1.0%	
							#	%	#	%
Levels	168	−0.03	−0.05	0.15	−0.34	0.62	3	1.79	0	0.00
Growth rates	142	−0.16	−0.15	0.21	−0.68	0.48	7	4.93	0	0.00
Ratios of two	11,929	−0.02	−0.02	0.17	−0.78	0.77	103	0.86	0	0.00
Ratios of three	2,373,539	−0.03	−0.03	0.17	−1.07	0.99	19,050	0.80	4	0.00

	Average return t-statistic									
	N	Mean	Median	Std	Min	Max	$\|t_\mu\| > 1.96$		$\|t_\mu\| > 2.57$	
							#	%	#	%
Levels	168	−0.31	−0.33	0.87	−2.69	2.46	10	5.95	3	1.79
Growth rates	142	−1.07	−1.08	1.35	−4.14	3.58	43	$\boxed{30.28}$	23	$\boxed{16.20}$
Ratios of two	11,929	−0.10	−0.13	0.98	−4.31	3.77	552	4.63	118	0.99
Ratios of three	2,373,539	−0.16	−0.18	0.98	−5.41	5.26	119,883	$\boxed{5.05}$	24,879	$\boxed{1.05}$

	Sharpe ratio									
	N	Mean	Median	Std	Min	Max	|SR| > 0.116		|SR| > 0.232	
							#	%	#	%
Levels	168	−0.01	−0.01	0.04	−0.12	0.11	2	1.19	0	0.00
Growth rates	142	−0.05	−0.05	0.06	−0.18	0.19	22	$\boxed{15.49}$	0	0.00
Ratios of two	11,929	−0.00	−0.01	0.04	−0.19	0.17	148	1.24	0	0.00
Ratios of three	2,373,539	−0.01	−0.01	0.04	−0.24	0.23	26,900	1.13	1	0.00

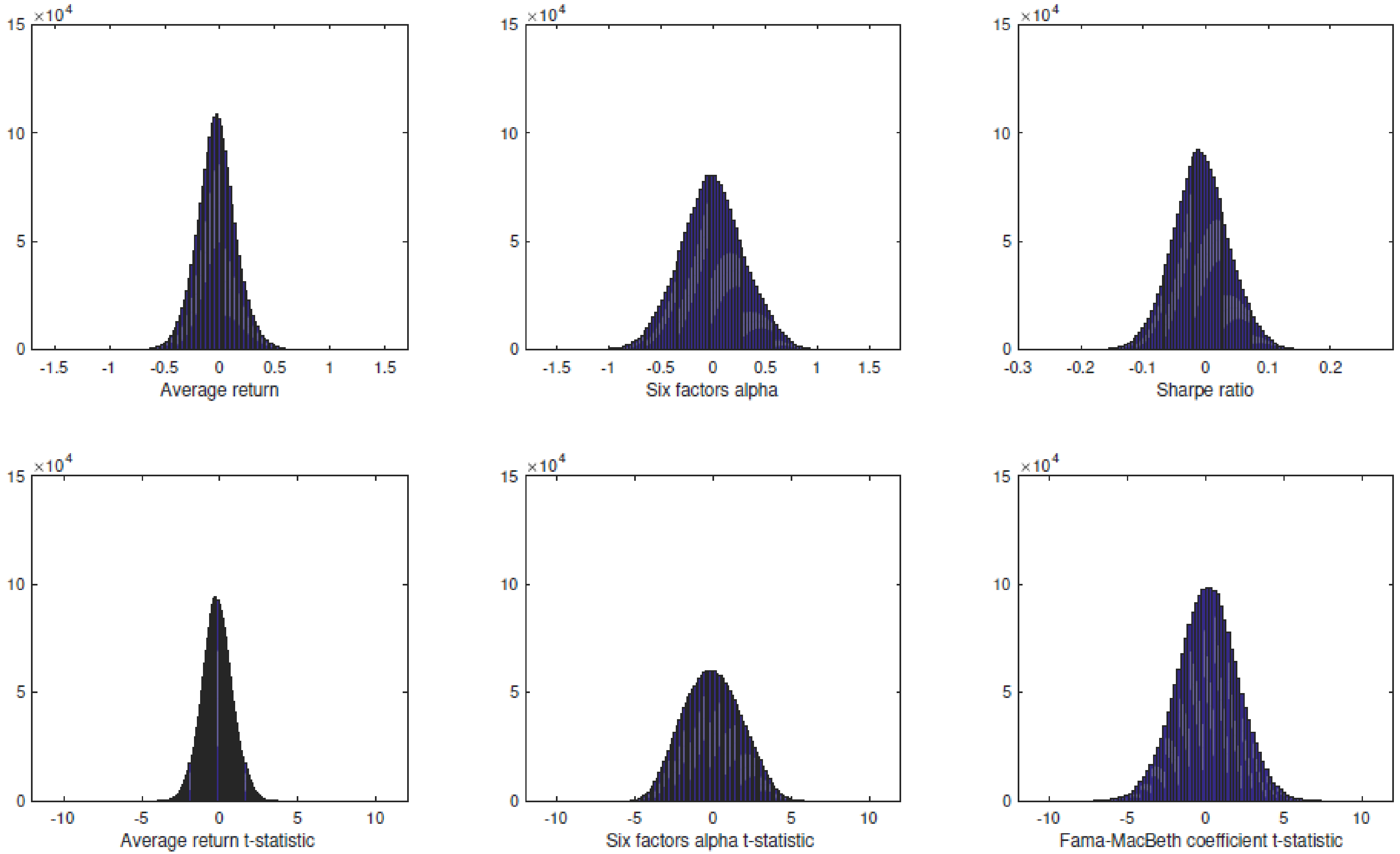
Summary statistics

	Mean	Median	Std	Min	Max	$\|t\| > 1.96$		$\|t\| > 2.57$	
						#	%	#	%
Panel A: Few strategies, Stocks filtered by size and price									
Alpha t-statistics									
CAPM	−0.08	−0.05	1.37	−5.46	4.15	1,882	15.38	780	6.37
FF3	−0.36	−0.39	1.54	−5.30	4.86	2,701	22.07	1,195	9.76
FF6	−0.58	−0.65	1.71	−4.96	5.71	3,621	29.59	2,012	16.44
BS	−0.79	−0.84	2.27	−6.51	7.43	6,122	44.16	4,316	31.13
HXZ	−0.56	−0.60	1.68	−5.05	5.02	4,089	29.50	2,215	15.98
Fama-MacBeth t-statistics									
CAPM	0.24	0.33	1.81	−7.33	5.79	3,502	28.61	1,948	15.92
FF3	0.31	0.34	1.30	−6.45	5.45	1,631	13.33	696	5.69
FF6	0.34	0.38	1.31	−6.40	5.78	1,671	13.65	749	6.12
BS	0.34	0.36	1.33	−6.66	6.25	1,935	14.00	894	6.47
HXZ	0.30	0.33	1.25	−6.69	5.62	1,696	12.27	683	4.94

Summary statistics

	Mean	Median	Std	Min	Max	$\|t\| > 1.96$		$\|t\| > 2.57$	
						#	%	#	%
Panel C: All strategies, Stocks filtered by size and price									
Alpha t-statistics									
CAPM	−0.37	−0.45	1.40	−5.74	6.78	434,302	18.20	163,436	6.85
FF3	−0.41	−0.44	1.46	−6.09	6.85	485,426	20.34	213,577	8.95
FF6	−0.05	−0.08	1.82	−6.75	7.36	724,442	30.36	401,271	16.82
BS	−0.09	−0.12	2.41	−7.94	7.73	1,085,859	45.51	760,742	31.88
HXZ	−0.15	−0.14	1.72	−6.33	6.51	659,498	27.64	342,001	14.33
Fama-MacBeth t-statistics									
CAPM	0.04	0.06	1.82	−7.55	6.93	686,718	28.78	382,160	16.02
FF3	−0.03	−0.02	1.56	−7.68	7.81	478,836	20.07	248,639	10.42
FF6	0.17	0.18	2.25	−11.68	10.93	938,357	39.33	623,755	26.15
BS	0.17	0.18	2.43	−11.27	11.35	999,646	41.90	693,572	29.07
HXZ	0.17	0.18	2.42	−12.50	11.06	997,478	41.81	693,143	29.05

Empirical Distributions



Cross-correlation

Problem

- Strategies rely on variables that are related
- There is cross-correlation in returns and residuals

Solution

- Implement a bootstrap
  - Kosowski, Timmermann, Wermers, and White (2006), Fama and French (2010), Yan and Zheng (2016)
  - Impose the null of alpha (or FM-delta) equal to zero
  - Bootstrap with replacement all returns and factors simultaneously to preserve cross-correlation
  - Stationary bootstrap – draw random blocks of 6 months
  - Similarly draw FM coefficients after subtracting the mean
  - Tabulate the statistics in 1,000 bootstrap iterations

Bootstrap results

	t_α		t_λ	
Percentile	Data	% Boot	Data	% Boot
0.5	−4.15	0.00	−5.49	0.00
1.0	−3.85	0.00	−4.97	0.00
2.5	−3.38	0.00	−4.20	0.00
5.0	−2.94	0.00	−3.52	0.00
10.0	−2.38	0.00	−2.72	0.00
20.0	−1.63	0.00	−1.71	0.10
30.0	−1.05	0.10	−0.99	0.60
40.0	−0.55	2.50	−0.39	6.20
50.0	−0.08	26.50	0.17	0.00
60.0	0.41	15.90	0.73	0.00
70.0	0.92	2.20	1.34	0.00
80.0	1.53	0.10	2.05	0.00
90.0	2.36	0.00	3.01	0.00
95.0	3.00	0.00	3.77	0.00
97.5	3.50	0.00	4.44	0.00
99.0	4.03	0.00	5.27	0.00
99.5	4.36	0.00	5.81	0.00

Multiple hypothesis testing

If one tests many null hypothesis, even if the experiments are **independent**, they cannot be evaluated using classical cutoffs

- Multiple testing means that some of them will be rejected by luck even if null is true

Example:

- Type I error for one test = $1 - 0.95 = 5\%$
- Type I error for ten tests = $1 - 0.95^{10} = 40\%$
- Type I error for 100 tests = $1 - 0.95^{100} = 99\%$

MHT approaches

Family-wise error rate, FWER

- FWER = Prob(Reject even one true null hypothesis)
- Control FWER ≤ a (significance level)

False discovery proportion, FDP

- FDP = #False rejections / #Total rejections
- Specify tolerance for FDP, γ (say 5%)
- Control Probability(FDP ≥ γ) ≤ α (say 5%)
- If FDP tolerance is 5%, we are willing to accept that at most 5% of ‘discovered’ anomalies may not be ‘real’

False discovery rate, FDR

- FDR = E(FDP)
- Control FDR ≤ g
  - No significance level

MHT approaches

FWER

- Bonferroni: Independence assumption
- Holm: Independence assumption
- StepM: Arbitrary cross-correlation (based on bootstrap, same as Harvey and Liu, 2016)

FDR

- BH: Limited cross-correlation
- BHY: Even more limited cross-correlation

FDP

- FDP-StepM

FWER

Bonferroni

- Reject null hypothesis, H_m at level α if $p_m \leq \frac{\alpha}{M}$
- M represents number of strategies being tested

Holm

- Rank p-values $p_1 \leq p_2 \leq \ldots \leq p_M$
- Reject H_i at level α if $p_i \leq \frac{\alpha}{(M-i+1)}$ for i=1 ... M

FWER

StepM method: Romano and Wolf (2005)

- Bootstrap data while maintaining correlations
- For each bootstrap iteration compute maximum t-stat
- $t_{max}^{(1)}, t_{max}^{(2)}, \ldots, t_{max}^{(B)},$ where B=1000
- Critical value c_1 is the (1- α) empirical percentile
- If for M_1 strategies, $t_m \geq c_1$ then M_1 are rejected
- Repeat bootstrap for M – M_1 strategies
- Repeat until no further strategies are rejected

FDR

BHY: Benjamini, Hochberg (1995), Benjamini, Yekutieli (2001)

- Rank p-values $p_1 \le p_2 \le \dots \le p_M$
- Reject null hypotheses, H_1, H_2, ..., H_{j*}

$$j^* = \max\left\{j : p_j \le \frac{j \times \delta}{M \times C_M}\right\} \qquad C_M = \sum_{i=1}^{M} 1/i \approx \log(M) + 0.5$$

δ Represents number of false rejections we are willing to tolerate

BH: Benjamini, Hochberg (1995)

$$j^* = \max\left\{j : p_j \le \frac{j \times \delta}{M}\right\}$$

FDP

Romano and Wolf (2007):

Control FDP at proportion γ and level α

$$\text{Prob}(FDP > \gamma) \leq \alpha$$

FDP-StepM method:

- Apply the k_j-StepM method

Monte Carlo Simulations

Return generating process

$$R_{pt} = \alpha_p + \beta_p' F_t + \epsilon_{pt}.$$

Draw factors and betas from MVN distributions with means and covariance matrix matched to cross-sectional distribution in data. Diagonal values of residuals also from MVN with mean zero and standard deviation matched to empirical distribution.

T = 500

Simulations = 1000

Strategies, N=10,000

Bootstraps = 1,000

Choose fraction f of non-zero alphas and correlation between the residuals.

Simulations: Yan and Zheng (2017)

f	0	0	0	5%	5%	5%	5%	5%	5%
α	0	0	0	0.5%	0.5%	0.5%	1.0%	1.0%	1.0%
ρ	0	3%	6%	0	3%	6%	0	3%	6%
Prct.	Frequency of bootstrap above actual t-statistic								
65	49.19	49.41	49.69	5.70	9.06	14.63	5.15	8.44	14.00
66	49.22	49.35	49.72	5.52	8.82	14.31	4.92	8.15	13.65
67	49.01	49.31	49.78	5.25	8.45	14.02	4.68	7.79	13.34
68	49.05	49.36	49.82	4.99	8.13	13.65	4.41	7.46	12.96
69	49.19	49.43	49.78	4.67	7.77	13.28	4.11	7.11	12.57
70	49.22	49.36	49.85	4.39	7.46	12.87	3.82	6.75	12.15
71	49.28	49.21	49.90	4.07	7.16	12.53	3.50	6.46	11.80
72	49.38	49.23	49.88	3.78	6.83	12.16	3.23	6.12	11.40
73	49.31	49.24	49.76	3.48	6.47	11.75	2.93	5.73	10.95
74	49.18	49.36	49.71	3.22	6.07	11.37	2.67	5.33	10.54
75	48.94	49.44	49.83	2.94	5.69	10.88	2.39	4.95	10.01
76	48.92	49.44	49.79	2.66	5.31	10.37	2.13	4.56	9.51
77	48.81	49.40	49.80	2.41	4.89	9.88	1.90	4.15	8.99
78	48.83	49.35	49.85	2.12	4.46	9.38	1.64	3.75	8.46
79	48.82	49.19	49.71	1.86	4.09	8.84	1.38	3.38	7.92
80	48.71	49.13	49.67	1.60	3.71	8.27	1.17	3.00	7.33
81	48.82	49.03	49.67	1.36	3.31	7.76	0.95	2.64	6.79
82	48.71	48.97	49.61	1.13	2.95	7.16	0.75	2.29	6.21
83	48.63	49.00	49.47	0.92	2.57	6.58	0.58	1.92	5.61
84	48.63	49.00	49.33	0.73	2.18	5.96	0.43	1.58	5.01
85	48.65	48.92	49.45	0.55	1.82	5.33	0.30	1.25	4.38
86	48.61	48.88	49.41	0.39	1.47	4.71	0.19	0.96	3.76
87	48.57	48.84	49.36	0.27	1.14	4.09	0.12	0.70	3.14
88	48.51	48.84	49.36	0.17	0.86	3.46	0.06	0.47	2.54
89	48.58	48.65	49.21	0.10	0.60	2.82	0.02	0.29	1.93
90	48.43	48.59	49.14	0.05	0.39	2.19	0.01	0.16	1.36
91	48.30	48.54	49.09	0.02	0.22	1.58	0.00	0.07	0.84
92	48.30	48.82	49.17	0.00	0.11	1.03	0.00	0.02	0.43
93	48.23	48.68	49.17	0.00	0.04	0.57	0.00	0.00	0.14
94	48.05	48.58	49.11	0.00	0.01	0.23	0.00	0.00	0.01
95	47.94	48.34	49.02	0.00	0.00	0.05	0.00	0.00	0.00
96	47.73	48.17	48.70	0.00	0.00	0.00	0.00	0.00	0.00
97	47.69	48.12	48.73	0.00	0.00	0.00	0.00	0.00	0.00
98	47.30	47.71	48.20	0.00	0.00	0.00	0.00	0.00	0.00
99	47.75	47.62	48.40	0.00	0.00	0.00	0.00	0.00	0.00

Simulations

Panel A: Basic properties (N = 10,000)							
			FWER		FDR		FDP
f	α	ρ	Bonf	Holm	BH	BHY	StepM
			Thresholds				
0	0.0	0	3.88	3.88	3.88	3.88	3.87
0	0.0	3	3.87	3.87	3.88	3.87	3.86
0	0.0	6	3.85	3.85	3.86	3.85	3.84
5	0.5	0	4.44	4.43	3.14	3.83	3.46
5	0.5	3	4.44	4.43	3.14	3.83	3.47
5	0.5	6	4.44	4.43	3.14	3.83	3.50
5	1.0	0	4.46	4.45	3.03	3.71	3.26
5	1.0	3	4.46	4.45	3.04	3.71	3.27
5	1.0	6	4.46	4.45	3.04	3.71	3.30
			Rejections rates				
0	0.0	0	0.00	0.00	0.00	0.00	0.00
0	0.0	3	0.00	0.00	0.00	0.00	0.00
0	0.0	6	0.00	0.00	0.00	0.00	0.00
5	0.5	0	1.76	1.77	3.45	2.44	2.94
5	0.5	3	1.76	1.77	3.45	2.44	2.92
5	0.5	6	1.77	1.77	3.46	2.44	2.87
5	1.0	0	4.59	4.60	5.19	4.86	5.04
5	1.0	3	4.59	4.60	5.20	4.86	5.03
5	1.0	6	4.59	4.60	5.19	4.86	5.01

Simulations

	FWER		FDR		FDP
	Bonf	Holm	BH	BHY	StepM
Panel B: Adaptive properties by varying f ($N = 10{,}000, \rho = 0, \alpha = 0.5\%$)					
f	Thresholds				
5	4.44	4.43	3.14	3.83	3.46
10	4.43	4.42	3.63	2.87	3.15
15	4.42	4.41	2.76	3.50	2.96
25	4.42	4.40	2.57	3.34	2.70
50	4.42	4.38	2.30	3.12	2.21
	Rejection rates				
5	1.76	1.77	3.45	2.44	2.94
10	3.54	3.56	5.37	7.61	6.69
15	5.33	5.36	11.71	8.51	10.77
25	8.89	8.99	20.49	15.20	19.57
50	17.81	18.24	43.20	33.15	44.23

Simulations

	FWER		FDR		FDP
	Bonf	Holm	BH	BHY	StepM
Panel B: Adaptive properties by varying f ($N = 10{,}000, \rho = 0, \alpha = 0.5\%$)					
Panel C: Adaptive properties by varying N ($f = 5\%, \rho = 0, \alpha = 0.5\%$)					
N	Thresholds				
1,000	4.14	4.14	3.19	3.84	3.63
10,000	4.44	4.43	3.14	3.83	3.46
50,000	4.89	4.89	3.13	3.88	3.38
100,000	5.03	5.03	3.13	3.90	3.37
500,000	5.33	5.33	3.13	3.93	3.37
1,000,000	5.45	5.45	3.13	3.94	3.37
	Rejections rates				
1,000	2.08	2.09	3.41	2.44	2.71
10,000	1.76	1.77	3.45	2.44	2.94
50,000	1.37	1.37	3.47	2.38	3.06
100,000	1.27	1.27	3.46	2.36	3.07
500,000	1.08	1.08	3.47	2.32	3.08
1,000,000	1.00	1.00	3.47	2.31	3.08

MHT critical values

	FWER				FDR				FDP	
	Bonferroni		Holm		BH		BHY		StepM	
	Thresh	%	Thresh	%	Thresh	%	Thresh	%	Thresh	%
Panel A: Few strategies, Stocks filtered by size and price										
Alpha t-statistic										
CAPM	4.61	0.04	4.90	0.04	3.47	1.04	4.90	0.04	4.53	0.06
FF3	4.61	0.14	4.61	0.14	3.02	5.08	4.27	0.41	3.54	4.03
FF6	4.61	0.20	4.63	0.20	2.69	14.45	3.89	1.99	3.91	1.94
BS	4.61	5.04	4.60	5.11	2.35	37.48	3.30	19.33	2.81	28.14
HXZ	4.61	0.24	4.61	0.24	2.68	14.83	3.96	1.47	4.73	0.11
Fama-MacBeth t-statistic										
CAPM	4.61	1.16	4.61	1.17	2.70	13.80	3.72	3.95	3.32	6.46
FF3	4.61	0.20	4.62	0.20	3.31	1.86	4.25	0.44	4.34	0.35
FF6	4.61	0.27	4.61	0.27	3.25	2.34	4.20	0.53	4.34	0.42
BS	4.61	0.40	4.61	0.40	3.19	2.88	4.11	0.85	4.04	0.88
HXZ	4.61	0.11	4.61	0.11	3.40	1.36	4.60	0.14	4.61	0.12

MHT critical values

	FWER				FDR				FDP	
	Bonferroni		Holm		BH		BHY		StepM	
	Thresh	%	Thresh	%	Thresh	%	Thresh	%	Thresh	%
Panel C: All strategies, Stocks filtered by size and price										
Alpha t-statistic										
CAPM	5.62	0.00	5.62	0.00	3.66	0.50	6.04	0.00	4.35	0.05
FF3	5.62	0.00	5.61	0.00	3.12	3.58	4.83	0.04	4.28	0.23
FF6	5.62	0.02	5.60	0.02	2.69	14.39	4.05	1.57	3.79	2.67
BS	5.62	0.92	5.60	0.92	2.36	36.41	3.47	15.92	2.76	27.61
HXZ	5.62	0.00	5.61	0.00	2.78	10.99	4.27	0.58	3.85	1.65
Fama-MacBeth t-statistic										
CAPM	5.62	0.08	5.62	0.09	2.72	13.15	3.92	2.76	4.11	2.06
FF3	5.62	0.06	5.62	0.06	2.96	6.21	4.10	1.25	3.95	1.60
FF6	5.62	1.06	5.61	1.06	2.47	27.15	3.57	10.75	3.12	16.31
BS	5.62	2.21	5.60	2.21	2.41	32.22	3.48	15.50	3.08	19.95
HXZ	5.62	2.04	5.60	2.04	2.41	32.17	3.48	15.29	3.08	19.88

Economic sanity checks

1. Signal has to cross both alpha and FM statistical thresholds
   - Intersection of the two sets
     - Drastically reduces the number of candidate

2. Signal has to satisfy economic threshold of Sharpe ratio at least as big as that of the market
   - Market Sharpe ratio, SRM = 0.116 (0.4 annually)

Proportion of Lucky Rejections

	Alpha	FM	Both Alpha and FM				
			All	0 to SRM/2	SRM/2 to SRM	SRM to 1.5×SRM	More than 1.5×SRM
Panel A: Few strategies, Stocks filtered by size and price							
Number of rejections by classical hypothesis testing							
CAPM	1,882	3,502	864	333	468	58	5
FF3	2,701	1,631	371	200	136	32	3
FF6	3,621	1,671	568	514	39	15	0
BS	5,652	1,760	961	827	113	21	0
HXZ	3,777	1,551	537	474	45	17	1
Proportion of lucky rejections after controlling FDR-BH							
CAPM	0.93	0.52	0.94	1.00	0.94	0.62	0.20
FF3	0.77	0.86	0.98	0.99	0.97	0.94	0.67
FF6	0.51	0.83	0.88	0.87	0.92	1.00	—
BS	0.19	0.80	0.77	0.74	0.93	1.00	—
HXZ	0.52	0.89	0.94	0.93	0.96	1.00	1.00
Proportion of lucky rejections after controlling FDP-StepM							
CAPM	1.00	0.77	1.00	1.00	1.00	1.00	0.20
FF3	0.91	0.97	1.00	1.00	1.00	1.00	1.00
FF6	0.93	0.97	0.99	0.99	1.00	1.00	—
BS	0.39	0.94	0.93	0.93	0.98	1.00	—
HXZ	1.00	0.99	1.00	1.00	1.00	1.00	1.00

Proportion of Lucky Rejections

	Alpha	FM	Both Alpha and FM				
			All	0 to SRM/2	SRM/2 to SRM	SRM to 1.5×SRM	More than 1.5×SRM
Panel C: All strategies, Stocks filtered by size and price							
Number of rejections by classical hypothesis testing							
CAPM	434,302	686,718	184,325	79,855	94,293	9,988	189
FF3	485,426	478,836	108,533	48,189	52,522	7,646	176
FF6	724,442	938,357	300,275	253,787	43,130	3,217	141
BS	1,085,859	999,646	476,018	403,408	68,984	3,507	119
HXZ	659,498	997,478	285,210	238,656	42,866	3,534	154
Proportion of lucky rejections after controlling FDR-BH							
CAPM	0.97	0.53	0.98	1.00	0.97	0.85	0.45
FF3	0.82	0.67	0.94	0.97	0.93	0.84	0.56
FF6	0.51	0.28	0.65	0.65	0.64	0.63	0.59
BS	0.20	0.23	0.37	0.36	0.40	0.45	0.35
HXZ	0.60	0.23	0.70	0.71	0.65	0.60	0.63
Proportion of lucky rejections after controlling FDP-StepM							
CAPM	1.00	0.92	1.00	1.00	1.00	0.99	0.84
FF3	0.99	0.92	1.00	1.00	1.00	0.99	0.84
FF6	0.91	0.59	0.96	0.97	0.94	0.93	0.96
BS	0.38	0.50	0.68	0.67	0.71	0.75	0.69
HXZ	0.94	0.51	0.98	0.98	0.94	0.94	0.97

Cast of survivors (17 out of 2.1 million)

(cstk − reajo) / xad	(Common-Ordinary Stock (Capital) − Retained Earnings Other Adjustments) / Advertising Expense
(lo − sppe) / tstkn	(Liabilities Other Total − Sale of Property) / Treasury Stock Number of Common Shares
(ap − txfed) / dvc	(Accounts Payable Trade − Income Taxes Federal) / Dividends Common-Ordinary
(csho − xsga) / xint	(Common Shares Outstanding − Selling, General and Administrative Expense) / Interest and Related Expense Total
(cshpri − xsga) / dd3	(Common Shares Used to Calculate Earnings Per Share Basic − Selling, General and Administrative Expense) / Debt Due in 3rd Year
(cshpri − xsga) / xint	(Common Shares Used to Calculate Earnings Per Share Basic − Selling, General and Administrative Expense) / Interest and Related Expense Total
(dcvsub − xrent) / dd2	(Debt Subordinated Convertible − Rental Expense) / Debt Due in 2nd Year
(dcvt − mrc5) / dltt	(Debt Convertible − Rental Commitments Minimum 5th Year) / Long-Term Debt Total
(dltis − pstkr) / mrc1	(Long-Term Debt Issuance − Preferred-Preference Stock Redeemable) / Rental Commitments Minimum 1st Year
(dltis − pstkr) / mrc2	(Long-Term Debt Issuance − Preferred-Preference Stock Redeemable) / Rental Commitments Minimum 2nd Year
(dltis − pstkr) / mrc3	(Long-Term Debt Issuance − Preferred-Preference Stock Redeemable) / Rental Commitments Minimum 3rd Year
(dltis − pstkr) / mrc4	(Long-Term Debt Issuance − Preferred-Preference Stock Redeemable) / Rental Commitments Minimum 4th Year
(dltis − pstkr) / mrct	(Long-Term Debt Issuance − Preferred-Preference Stock Redeemable) / Rental Commitments Minimum 5 Year Total
(rectr − xsga) / xint	(Receivables Trade − Selling, General and Administrative Expense) / Interest and Related Expense Total
(esubc − txdi) / dpvieb	(Equity in Net Loss Earnings − Income Taxes Deferred) / Depreciation (Accumulated) Ending Balance (Schedule VI)
(txdi − xpr) / dpvieb	(Income Taxes Deferred − Pension and Retirement Expense) / Depreciation (Accumulated) Ending Balance (Schedule VI)
(pstkc − txdi) / ppeveb	(Preferred Stock Convertible − Income Taxes Deferred) / Property, Plant, and Equipment Ending Balance (Schedule V)

Cast of Survivors (17 out of 2.1 million)

	Panel A: Descriptive statistics					
	Mean	t_μ	SR	Alpha	t_α	t_λ
(cstk − reajo) / xad	−0.67	−2.33	−0.12	−1.20	−4.37	−3.55
(lo − sppe) / tstkn	0.40	3.00	0.13	0.55	3.97	3.22
(ap − txfed) / dvc	−0.49	−2.99	−0.13	−0.61	−3.82	−3.54
(csho − xsga) / xint	−0.77	−3.44	−0.15	−0.95	−3.96	−4.82
(cshpri − xsga) / dd3	−0.66	−3.19	−0.15	−0.87	−3.95	−4.02
(cshpri − xsga) / xint	−0.64	−2.78	−0.12	−1.01	−4.22	−4.84
(dcvsub − xrent) / dd2	−0.49	−3.32	−0.15	−0.71	−4.67	−3.16
(dcvt − mrc5) / dltt	−0.44	−2.99	−0.14	−0.58	−3.88	−3.66
(dltis − pstkr) / mrc1	−0.48	−2.64	−0.13	−0.85	−4.58	−3.12
(dltis − pstkr) / mrc2	−0.47	−2.57	−0.13	−0.85	−4.38	−3.96
(dltis − pstkr) / mrc3	−0.51	−2.77	−0.14	−0.89	−4.58	−4.21
(dltis − pstkr) / mrc4	−0.57	−3.04	−0.15	−0.91	−4.46	−3.46
(dltis − pstkr) / mrct	−0.50	−2.81	−0.14	−0.92	−5.12	−3.57
(rectr − xsga) / xint	−0.60	−2.82	−0.13	−1.04	−4.90	−3.60
(esubc − txdi) / dpvieb	0.64	3.45	0.15	1.08	6.36	3.94
(txdi − xpr) / dpvieb	−0.45	−2.86	−0.13	−0.68	−3.97	−4.98
(pstkc − txdi) / ppeveb	0.38	2.75	0.12	0.67	4.93	3.86

Bayesianized p-values

Harvey (2017)

- Symmetric and Descending minimum Bayes factor
- sdMBF = -exp(1) * p-value * log(p-value)
- Posterior Bayesianized p-value is

$$\frac{sdMBF \; * Prior\ Odds}{1 + sdMBF \; * Prior\ Odds}$$

Bayesianized p-values

	Panel B: Bayesian p-values											
	Posterior p-value						Prior odds ratio					
	t_α			t_λ			t_α			t_λ		
	Prior odds ratio			Prior odds ratio			Posterior p-value			Posterior p-value		
	99 to 1	95 to 5	90 to 10	99 to 1	95 to 5	90 to 10	0.01	0.05	0.10	0.01	0.05	0.10
(cstk − reajo) / xad	0.036	0.007	0.003	0.446	0.134	0.068	0.964	0.993	0.997	0.554	0.866	0.932
(lo − sppe) / tstkn	0.156	0.034	0.017	0.700	0.309	0.175	0.844	0.966	0.983	0.300	0.691	0.825
(ap − txfed) / dvc	0.241	0.058	0.028	0.458	0.139	0.071	0.759	0.942	0.972	0.542	0.861	0.929
(csho − xsga) / xint	0.162	0.036	0.017	0.005	0.001	0.000	0.838	0.964	0.983	0.995	0.999	1.000
(cshpri − xsga) / dd3	0.165	0.037	0.018	0.133	0.029	0.014	0.835	0.963	0.982	0.867	0.971	0.986
(cshpri − xsga) / xint	0.064	0.013	0.006	0.005	0.001	0.000	0.936	0.987	0.994	0.995	0.999	1.000
(dcvsub − xrent) / dd2	0.010	0.002	0.001	0.731	0.343	0.198	0.990	0.998	0.999	0.269	0.657	0.802
(dcvt − mrc5) / dltt	0.202	0.046	0.023	0.360	0.097	0.049	0.798	0.954	0.977	0.640	0.903	0.951
(dltis − pstkr) / mrc1	0.015	0.003	0.001	0.753	0.369	0.217	0.985	0.997	0.999	0.247	0.631	0.783
(dltis − pstkr) / mrc2	0.035	0.007	0.003	0.162	0.036	0.017	0.965	0.993	0.997	0.838	0.964	0.983
(dltis − pstkr) / mrc3	0.015	0.003	0.001	0.067	0.014	0.006	0.985	0.997	0.999	0.933	0.986	0.994
(dltis − pstkr) / mrc4	0.025	0.005	0.002	0.521	0.173	0.090	0.975	0.995	0.998	0.479	0.827	0.910
(dltis − pstkr) / mrct	0.001	0.000	0.000	0.436	0.129	0.066	0.999	1.000	1.000	0.564	0.871	0.934
(rectr − xsga) / xint	0.004	0.001	0.000	0.408	0.117	0.059	0.996	0.999	1.000	0.592	0.883	0.941
(esubc − txdi) / dpvieb	0.000	0.000	0.000	0.172	0.038	0.019	1.000	1.000	1.000	0.828	0.962	0.981
(txdi − xpr) / dpvieb	0.157	0.034	0.017	0.002	0.000	0.000	0.843	0.966	0.983	0.998	1.000	1.000
(pstkc − txdi) / ppeveb	0.003	0.001	0.000	0.216	0.050	0.024	0.997	0.999	1.000	0.784	0.950	0.976

Conclusions

- The profession might be going down a dangerous road, which is a lot more slippery than what Harvey, Liu, and Zhu (2016) have warned us about

- If you believe our thought experiment, the *t*-stat threshold you should use is closer to 4 (at 5% significance level)

- If you believe that, none of the anomalies that people are talking about are significant

- The only strategies that would be significant appear to be totally nonsensical
  - Use theory to motivate strategies