Quantitative Asset Management

Bernard Herskovic

Spring 2025

1. Does Academic Research Destroy Stock Return Predictability?
   McLean and Pontiff (2016, JF)
2. ... and the Cross-Section of Expected Returns
   Harvey, Liu, and Zhu (2016, RFS)
3. Replicating Anomalies
   Hou, Xue, Zhang (2020, RFS)

Does Academic Research Destroy Stock Return Predictability?

McLean and Pontiff (2016, JF)

Does Academic Research Destroy Stock Return Predictability?

- Finance literature documented several cross-sectional relations between characteristics and future stock return
- Studies rely on specific sample available at the time of publication
- What happen after an anomaly is documented?

Does Academic Research Destroy Stock Return Predictability?

McLean and Pontiff (2016):

- Focus on 97 characteristics that predict return
  - Published results!
  - Only cross-sectional relations (no time series predictability)
- Look for results out-of-sample:
  - Original sample
  - Sample after study but before publication
  - Use post-publication sample

Does Academic Research Destroy Stock Return Predictability?

- Explanations for predictability results:
  - Statistical bias
  - Rational pricing
  - Mispricing
- What to expect from out-of-sample results?

Method

- Identify set of cross-sectional relations
  - Peer-reviewed finance, accounting, and economics journals
  - Uses publicly available data
  - Predictability identified via (i) Fama-MacBeth regressions and/or (ii) long-short portfolios
  - 79 studies and 97 cross-sectional relations
  - Long-short quintiles
  - Two important dates
    - Publication date (relation is public)
    - End-of-sample date (statistical bias)

Summary Statistics

Table 1

Number of predictor portfolios	97
Predictors portfolios with t-statistic > 1.5	85 (88%)
Mean publication year	2000
Median publication year	2001
Predictors from finance journals	68 (70%)
Predictors from accounting journals	27 (28%)
Predictors from economics journals	2 (2%)
Mean portfolio return in-sample	0.582
Standard deviation of mean in-sample portfolio return	0.395
Mean observations in-sample	323
Mean portfolio return out-of sample	0.402
Standard deviation of mean out-of-sample portfolio return	0.651
Mean observations out-of-sample	56
Mean portfolio return post-publication	0.264
Standard deviation of mean post-publication portfolio return	0.516
Mean observations post-publication	156

Empirical Analyses

Baseline regression:

$$R_{it} = \alpha_i + \beta_1 \, Post\ Sample\ Dummy_{i,t} + \beta_2 \, Post\ Publication\ Dummy_{i,t} + e_{it}$$

- β_1: ?
- β_2: ?
- Is this specification able to disentangle statistical bias and impact of publication?

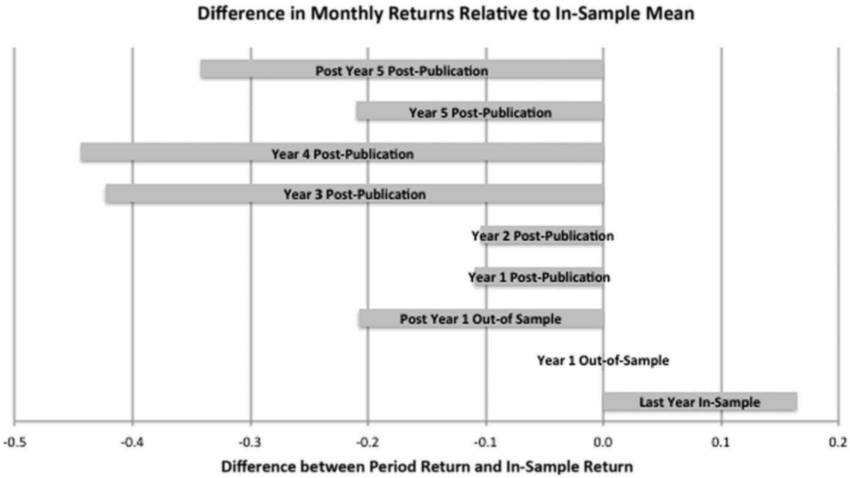
Results

Table 2

Variables	(1)	(2)	(3)	(4)
Post-Sample (S)	−0.150***	−0.180**	0.157	0.067
	(0.077)	(0.085)	(0.103)	(0.112)
Post-Publication (P)	−0.337***	−0.387***	−0.002	−0.120
	(0.090)	(0.097)	(0.078)	(0.114)
S × Mean			−0.532***	
			(0.221)	
P × Mean			−0.548***	
			(0.178)	
S × t-statistic				−0.061***
				(0.023)
P × t-statistic				−0.063***
				(0.018)
Predictor FE?	Yes	Yes	Yes	Yes
Observations	51,851	45,465	51,851	51,944
Predictors (N)	97	85	97	97
Null : S = P	0.024	0.021		
Null: P = −1 × (mean)	0.000	0.000		
Null: S = −1 × (mean)	0.000	0.000		

Finer post-sample and post-publication partitions

Figure 2



Time, persistence or publication effect?

- Time: costs?
- Persistence: Value and Momentum everywhere?
- Publication?

Time, persistence or publication effect?

Table 3

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Time	−0.069***		−0.069***			
	(0.011)		(0.026)			
Post-1993		−0.120	0.303***			
		(0.074)	(0.118)			
Post-Sample			−0.190**	−0.179**	−0.132*	−0.128
			(0.081)	(0.080)	(0.076)	(0.078)
Post-Publication			−0.362***	−0.310**	−0.295***	−0.258***
			(0.124)	(0.122)	(0.089)	(0.093)
1-Month Return					0.114***	
					(0.015)	
12-Month Return						0.020***
						(0.004)
Observations	51,851	51,851	51,851	51,851	51,754	50,687
Char. FE?	Yes	Yes	Yes	Yes	Yes	Yes
Time FE?	No	No	No	Yes	No	No

Decay varying across types?

- Four broad categories
  - Event
  - Market
  - Valuation
  - Fundamentals
- Test:

$$R_{i,t} = \alpha_{i,} + \beta_1 \, Post\ Publication\ Dummy_i + \beta_2 \, Predictor\ Type\ Dummy_i$$
$$+ \beta_3 \, Post\ Publication\ Dummy_i \times Predictor\ Type\ Dummy_i + e_{it}$$

What does each coefficient mean?

Decay varying across types?

Table 4

Variable	(1)	(2)	(3)	(4)
Post-Publication (P)	−0.208***	−0.316***	−0.310***	−0.301***
	(0.059)	(0.097)	(0.080)	(0.089)
Market	0.304***			
	(0.079)			
P × Market	−0.244			
	(0.169)			
Event		−0.098**		
		(0.046)		
P × Event		0.105		
		(0.091)		
Valuation			−0.056	
			(0.063)	
P × Valuation			0.186	
			(0.131)	
Fundamental				−0.201***
				(0.045)
P × Fundamental				0.025
				(0.089)
Constant	0.482***	0.606***	0.585***	0.630***
	(0.036)	(0.052)	(0.000)	(0.053)
Observations	51,851	51,851	51,851	51,851
Predictors	97	97	97	97
Type + (P × Type)	0.060	0.007	0.121	−0.176
p-value	0.210	0.922	0.256	0.012

Arbitrage costs

- Arbitrage cost may explain mispricing
- What are the post-publication effects when arbitrage is costlier?
- Examples?

Trading activity

Table 6

Variables	Variance	Trading volume	Dollar volume	Short-long short interest
Post-Sample (S)	−0.054***	0.092***	0.066***	0.166***
	(0.007)	(0.001)	(0.007)	(0.014)
Post-Publication (P)	−0.065***	0.187***	0.097***	0.315***
	(0.008	(0.013)	(0.007)	(0.013)
Observations	52,632	52,632	52,632	41,026
Time FE?	Yes	Yes	Yes	No
Predictor FE?	Yes	Yes	Yes	Yes
Null: S = P	0.156	0.000	0.000	0.000

Summary

- Average predictor's return decline 58%
- Statistical bias: 26%
- Publication effect: 32%
- Trading activity

... and the Cross-Section of Expected Returns

Harvey, Liu, and Zhu (2016, RFS)

... and the Cross-Section of Expected Returns

- Many years of research in asset pricing
- Several ‘factors’ and ‘anomalies’ have been identified
- How to test them? t-statistics!
  - When to reject the null?

“Given the known number of factor that have been tried and the reasonable assumption that many more factors have been tried but did not make it to publication, the usual cutoff levels for statistical significance may not be appropriate”

Hypothesis testing

- Compute t-statistic
- Under the null, how likely should be to observe the estimated statistic?
- p-value: probability of observing a statistic equal or more extreme under the null assumption
  - We reject the null when p-value is low
  - We cannot reject the null when p-value is high
- error type 1
  - We reject the null although the null is true
- error type 2
  - We do not reject the null although the null is false
- By focusing on p-value, we try to avoid type 1 error

Limitations

- Factors are not all the same
  - Economic mechanism: theory
    vs.
  - Pure empirical exercises
- Focus on unconditional moments
  - Time-varying prices of risk
  - Consider conditional models
  - Maybe the market price of risk is higher when dividend yield is higher...
- Focus on the cross-section of returns

... and the Cross-Section of Expected Returns

- Which factors to include?
  - Include different empirical proxies for the same type of risk
  - Factors that explain return pattern in general
  - Top accounting, finance, and economics journal
  - Include working papers too
- 313 articles:
  - 250 published
  - 63 working papers
  - 316 different factors!

Factor classification

Table 1

Risk type		Description	Examples
Common (113)	**Financial** (46)	Proxy for aggregate financial market movement, including market portfolio returns, volatility, squared market returns, among others	Sharpe (1964): market returns; Kraus and Litzenberger (1976): squared market returns
	Macro (40)	Proxy for movement in macroeconomic fundamentals, including consumption, investment, inflation, among others	Breeden (1979): consumption growth; Cochrane (1991): investment returns
	Microstructure (11)	Proxy for aggregate movements in market microstructure or financial market frictions, including liquidity, transaction costs, among others	Pastor and Stambaugh (2003): market liquidity; Lo and Wang (2006): market trading volume
	Behavioral (3)	Proxy for aggregate movements in investor behavior, sentiment or behavior-driven systematic mispricing	Baker and Wurgler (2006): investor sentiment; Hirshleifer and Jiang (2010): market mispricing
	Accounting (8)	Proxy for aggregate movement in firm-level accounting variables, including payout yield, cash flow, among others	Fama and French (1992): size and book-to-market; Da and Warachka (2009): cash flow
	Other (5)	Proxy for aggregate movements that do not fall into the above categories, including momentum, investors' beliefs, among others	Carhart (1997): return momentum; Ozoguz (2009): investors' beliefs
Characteristics (202)	**Financial** (61)	Proxy for firm-level idiosyncratic financial risks, including volatility, extreme returns, among others	Ang et al. (2006): idiosyncratic volatility; Bali, Cakici, and Whitelaw (2011): extreme stock returns
	Microstructure (28)	Proxy for firm-level financial market frictions, including short sale restrictions, transaction costs, among others	Jarrow (1980): short sale restrictions; Mayshar (1981): transaction costs
	Behavioral (3)	Proxy for firm-level behavioral biases, including analyst dispersion, media coverage, among others	Diether, Malloy, and Scherbina (2002): analyst dispersion; Fang and Peress (2009): media coverage
	Accounting (87)	Proxy for firm-level accounting variables, including PE ratio, debt-to-equity ratio, among others	Basu (1977): PE ratio; Bhandari (1988): debt-to-equity ratio
	Other (24)	Proxy for firm-level variables that do not fall into the above categories, including political campaign contributions, ranking-related firm intangibles, among others	Cooper, Gulen, and Ovtchinnikov (2010): political campaign contributions; Edmans (2011): intangibles

The numbers in parentheses represent the number of factors identified. See Table 6 and http://faculty.fuqua.duke.edu/∼charvey/Factor-List.xlsx.

... and the Cross-Section of Expected Returns

- Biases
  - It is hard to publish a nonresult (publication bias)
  - Always looking for new factors
  - Hard to publish replication exercises
- How to test for these biases?
  - Out-of-sample (e.g. McLean and Pontiff 2016)
  - Some strategies don't survive simple replications
  - What is a key issue with out-of-sample testing?
- Multiple testing framework

Example

Table 2

Panel A: An example

	Unpublished	Published	Total
Truly insignificant	500	50	550
Truly significant	100	50	150
Total	600	100(R)	700(M)

Panel B: The testing framework

	H_0 not rejected	H_0 rejected	Total
H_0 true	$N_{0\|a}$	$N_{0\|r}$	M_0
H_0 false	$N_{1\|a}$	$N_{1\|r}$	M_1
Total	$M-R$	R	M

Panel A shows a hypothetical example for factor testing. Panel B presents the corresponding notation in a standard multiple testing framework.

- 50 are false discoveries (error type 1)
- 100 true factors are not published (error type 2)
- Higher probability of false discoveries in published works

... and the Cross-Section of Expected Returns

- How should we account for the joint occurrence of false discoveries?

Family-wise error rate

- It is the probability of observing at least one false discovery

$$\text{FWER} = Pr(N_{0|r} \geq 1)$$

- α_w: significance level
- Goal: low FWER

False discovery rate

- Fraction of false discoveries:

$$\text{FDP} = \begin{cases} \dfrac{N_{0|r}}{R} & \text{if } R > 0, \\ 0 & \text{if } R = 0. \end{cases}$$

- False discovery rate:

$$\text{FDR} = E[FDP]$$

- α_d: significance level
- Goal: low FDR

Three approaches to adjust p-value

Table 3

Adjustment type	Single/Sequential	Multiple test
Bonferroni	single	family-wise error rate
Holm	sequential	family-wise error rate
Benjamini, Hochberg, and Yekutieli (BHY)	sequential	false discovery rate

Bonferroni

- Reject if

$$\text{p-value} \leq \frac{\alpha_w}{M}$$

which is the same as

$$p_i^{\text{Bonferroni}} = \min\{Mp_i, 1\}$$

Example

Table 4

Panel A: Single tests and "significant" factors

Test →	1	2	3	4	5	6	7	8	9	10	# of discoveries
t-statistic	1.99	2.63	2.21	3.43	2.17	2.64	4.56	5.34	2.75	2.49	10
p-value (%)	**4.66**	**0.85**	**2.71**	**0.05**	**3.00**	**0.84**	**0.00**	**0.00**	**0.60**	**1.28**	

Panel B: Bonferroni "significant" factors

Test →	1	2	3	4	5	6	7	8	9	10	
t-statistic	1.99	2.63	2.21	3.43	2.17	2.64	4.56	5.34	2.75	2.49	3
p-value (%)	4.66	0.85	2.71	**0.05**	3.00	0.84	**0.00**	**0.00**	0.60	1.28	

Hom's adjustment

- Sequential:
  1. Order p-values: $p_{(1)} \leq p_{(2)} \leq \cdots p_{(b)} \leq \cdots p_{(M)}$
  2. Set k to be the minimum index such that $p_{(b)} > \frac{\alpha_w}{M+1-b}$
  3. Reject all below k

$$p_{(i)}^{Holm} = \min[\max_{j \leq i}\{(M-j+1)p_{(j)}\}, 1]$$

Example

Table 4

Panel A: Single tests and "significant" factors

Test →	1	2	3	4	5	6	7	8	9	10	# of discoveries
t-statistic	1.99	2.63	2.21	3.43	2.17	2.64	4.56	5.34	2.75	2.49	10
p-value (%)	**4.66**	**0.85**	**2.71**	**0.05**	**3.00**	**0.84**	**0.00**	**0.00**	**0.60**	**1.28**	

Panel B: Bonferroni "significant" factors

Test →	1	2	3	4	5	6	7	8	9	10	
t-statistic	1.99	2.63	2.21	3.43	2.17	2.64	4.56	5.34	2.75	2.49	3
p-value (%)	4.66	0.85	2.71	**0.05**	3.00	0.84	**0.00**	**0.00**	0.60	1.28	

Panel C: Holm adjusted p-values and "significant" factors

Reordered tests b	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
Old order	8	7	4	9	6	2	10	3	5	1	4
p-value (%)	**0.00**	**0.00**	**0.05**	**0.60**	0.84	0.85	1.28	2.71	3.00	4.66	
$\alpha_w/(M+1-b)$ $\alpha_w=5\%$	0.50	0.56	0.63	0.71	0.83	1.00	1.25	1.67	2.50	5.00	

Benjamini, Hochberg, and Yekutieli's adjustment

- Sequential:
  1. Order p-values: $p_{(1)} \leq p_{(2)} \leq \cdots p_{(b)} \leq \cdots p_{(M)}$
  2. Set k to be the *maximum* index such that $p_{(b)} \leq \frac{b}{M \times c(M)} \alpha_d$
  3. Reject all below k

$$p_{(i)}^{BHY} = \begin{cases} p_{(M)} & \text{if } i = M, \\ \min[p_{(i+1)}^{BHY}, \frac{M \times c(M)}{i} p_{(i)}] & \text{if } i \leq M-1 \end{cases}$$

where

$$c(M) = \sum_{j=1}^{M} \frac{1}{j}$$

Example

Table 4

Panel A: Single tests and "significant" factors

Test →	1	2	3	4	5	6	7	8	9	10	# of discoveries
t-statistic	1.99	2.63	2.21	3.43	2.17	2.64	4.56	5.34	2.75	2.49	10
p-value (%)	**4.66**	**0.85**	**2.71**	**0.05**	**3.00**	**0.84**	**0.00**	**0.00**	**0.60**	**1.28**	

Panel B: Bonferroni "significant" factors

Test →	1	2	3	4	5	6	7	8	9	10	
t-statistic	1.99	2.63	2.21	3.43	2.17	2.64	4.56	5.34	2.75	2.49	3
p-value (%)	4.66	0.85	2.71	**0.05**	3.00	0.84	**0.00**	**0.00**	0.60	1.28	

Panel C: Holm adjusted p-values and "significant" factors

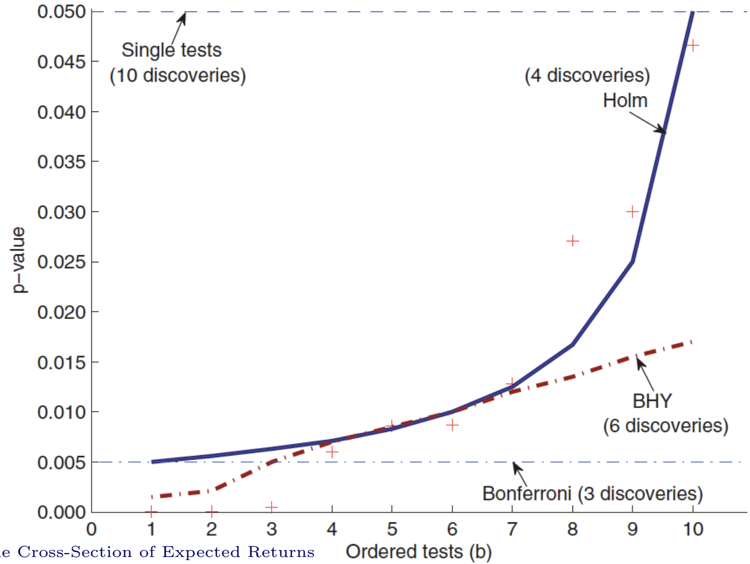
Reordered tests b	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
Old order	8	7	4	9	6	2	10	3	5	1	4
p-value (%)	**0.00**	**0.00**	**0.05**	**0.60**	0.84	0.85	1.28	2.71	3.00	4.66	
$\alpha_w/(M+1-b)$ $\alpha_w = 5\%$	0.50	0.56	0.63	0.71	0.83	1.00	1.25	1.67	2.50	5.00	

Panel D: BHY adjusted p-values and "significant" factors

Reordered tests b	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
Old order	8	7	4	9	6	2	10	3	5	1	6
p-value (%)	**0.00**	**0.00**	**0.05**	**0.60**	**0.84**	**0.85**	1.28	2.71	3.00	4.66	
$(b \cdot \alpha_d)/(M \times c(M))$ $\alpha_d = 5\%$	0.17	0.34	0.51	0.68	0.85	1.02	1.19	1.37	1.54	1.71	

Example

Figure 1



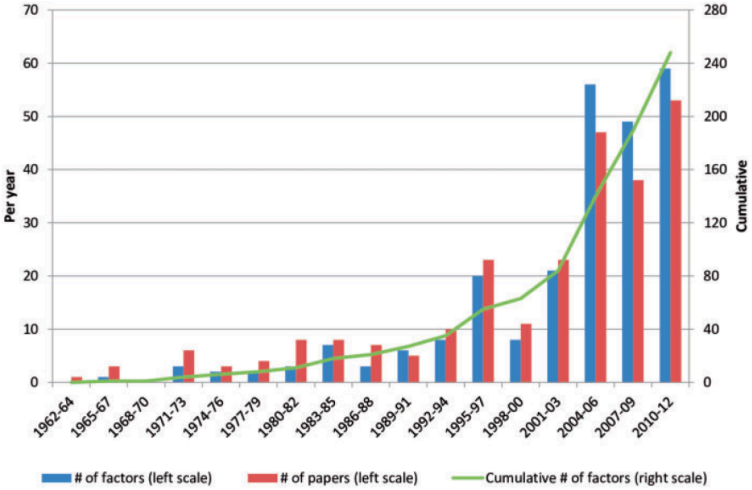
… and the Cross-Section of Expected Returns

Data

- 316 different factors
- Use 5% significance
- Assume all are ‘econometrically sound’
- Take published t-statistic as given

Factors discovery over time

Figure 2



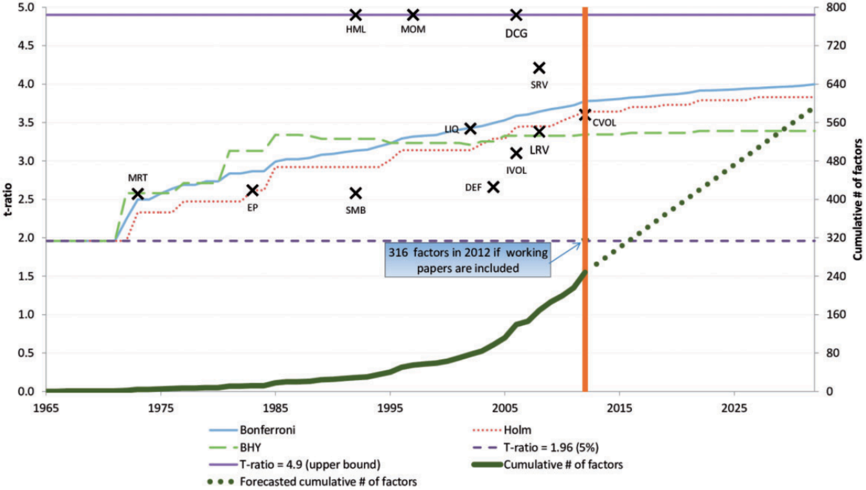
… and the Cross-Section of Expected Returns

p-value adjustment:

- Assume all tests are published: $M = R$
  - Lower bond
  - Factors are already an ‘elite’ group (they were published)
- Use $\alpha_w = 5\%$
- Use $\alpha_d = 1\%$

Adjusted p-value thresholds

Figure 3



... and the Cross-Section of Expected Returns

- t-stat of 3?
  - p-value of .27% seems a bit too high
  - but they only count published factors
  - only consider a small fraction of working papers
  - culture of discovery of new factors
  - 316 is only a small fraction of all factors documented
  - maybe a t-stat of 3 is not that high after all
- Out of 316 factors, 158 are arguably false

"We attempt to navigate the zoo and establish new benchmarks to guide empirical asset pricing test."

Replicating Anomalies

Hou, Xue, Zhang (2017, working paper)

Replicating Anomalies

"This paper conducts a gigantic replication of the entire anomalies literature compiling a largest-to-date data library with 447 anomaly variables."

Replicating Anomalies

447 anomalies in the following categories:

- 57 momentum
- 68 value-versus-growth
- 38 investment
- 79 profitability
- 103 intangibles
- 102 trading frictions

Data mining

"Out of 447 anomalies, 286 (64%) are insignificant at the 5% level. Imposing the cutoff t-value of three proposed by Harvey, Liu, and Zhu (2016) raises the number of insignificant anomalies further to 380 (85%)."

q-factor model

To explain the 161 significant anomalies, they use the q-factor model developed by the same authors

- The q-factor model is a four-factor model (Hou, Xue, and Zhang 2015 RFS):
  - Market
  - Size (SMB)
  - Investment-to-asset (I/A)
  - Return on equity (Roe)
- They don't use the five-factor model by Fama and French (2015):
  - Market
  - Size (SMB)
  - Book-to-market (HML)
  - Investment (CMA) — similar to I/A
  - Profitability (RMW) — similar to Roe
- Profitability and investment factors' construction similar to HML procedure

Replicating Anomalies

Replication procedure:

- Return data from CRSP
- Accounting information from Compustat
- Sample from January 1967 to December 2014
- Exclude financial firms
- Exclude firms with negative book value

Important replication assumptions

1. NYSE breakpoints
   - Large cross-sectional dispersion of anomalies among microcaps
2. Value-weighted returns
   - Capture wealth effect
   - Microcaps (bottom NYSE quintile) are influential in equal-weighted portfolios
   - Other trading frictions are less pronounced
3. No Fama-MacBeth estimation, only portfolio sorts
   - Microcaps: FM are zero-investment portfolios, more similar to equal-weighted
   - Microcaps: linear functional form may be affects by outliers (i.e. volatile returns, etc.)
   - FM portfolios require high turnover
4. Focus on high-minus-low decile

Results: Table 3

447 anomalies in the following categories:

- 57 momentum
  - 20 are not significant
- 68 value-versus-growth
  - 37 are not significant
- 38 investment
  - 11 are not significant
- 79 profitability
  - 46 are not significant
- 103 intangibles
  - 77 are not significant
- 102 trading frictions
  - 95 are not significant

What do we learn from this exercise?

What do we learn from this exercise?

- Lots of anomalies do not survive replication
- Replication did not follow the same rules as the original papers
- Replication did not estimate prices of risk
- Could estimate time-varying prices of risk
- Focus on one particular factor model (q-factor model)
- Although lots of anomalies become not significant, the main ones are still there:
  - Book-to-market
  - Momentum

What do we learn from this exercise?

Should we look for more anomalies?

What do we learn from this exercise?

"Perhaps most important, the credibility of the anomalies literature can improve via a closer connection with economic theory"