Empirical Asset Pricing

Francesco Franzoni

Swiss Finance Institute - USI Lugano

Cross-Sectional Anomalies: The Debate

Lecture Outline

1. Fama and French's results: size and B/M
2. Daniel and Titman reply to FF
3. Other anomalies: reversal, momentum, profitability and investment, idiosyncratic volatility, betting against beta
4. Kozak, Nagel, and Santosh: Intepreting Factor Models
5. Replication crisis

Relevant readings:

- Fama and French, 1992, "The Cross-Section of Expected Stock Returns", Journal of Finance
- Fama and French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds", Journal of Financial Economics

- Cochrane, section 20.2
- Daniel and Titman, 1997, "Evidence on the Characteristics of Cross-Sectional Variation in Stock Returns", Journal of Finance
- Kozak, Nagel, and Santosh, 2018, Intepreting Factor Models, Journal of Finance
- Ang, Hodrick, Xing, and Zhang, 2009, "High Idiosyncratic Volatility and Low Returns: International and Further U.S. Evidence", JFE
- Frazzini and Pedersen, 2014, "Betting against Beta", Journal of Financial Economics
- Harvey, Liu, and Zhu, 2016, . . . and the cross-section of expected returns, Review of Financial Studies
- Jensen, Kelly, and Pedersen, 2023, Is there a replication crisis in finance?, Journal of Finance

1. Fama and French: Size and B/M

Fama and French, 1992, Journal of Finance

- Prior work documented several deviations from CAPM: size (Banz, 1981), leverage (Bhandari, 1988), B/M (Stattman, 1980), E/P (Basu, 1983)

- These price ratios likely capture similar effects. Goal: identify which ones really matter

Results:

1. **Beta has no role** in explaining cross-section of returns in 1963-1990: *clear rejection of CAPM*

2. Size, E/P, B/M, Leverage are all individually significant in univariate tests. But in multivariate tests, **only Size and especially B/M are significant**

Data and Methodology

- All (non-financial) firms in the intersection of CRSP and Compustat: 1963-1990
- Match accounting data of December year $t-1$ to returns from July of year t to June of year $t+1$. Goal: allow some time for information to spread
  - More recent implementations develop more timely measures of B/M (e.g., Asness and Frazzini, 2013)
- **Fama and MacBeth (1973) regressions**: regress monthly returns on betas and accounting variables
- **Key methodological issue:** betas are measured with error at firm level, so they assign to each stock the beta of a size/beta-sorted portfolio to which it belongs (100 portfolios from 10×10 sort). Post-ranking betas are estimated on the full sample (Dimson adjustment)
- **Tradeoff:** portfolio grouping reduces measurement error but also reduces dispersion of betas and power of tests

Regression Results

- The cross-sectional regression they run on each month of data is

$$R_{it} = \lambda\beta_{it} + \gamma z_i + u_{it} \qquad i = 1...N$$

  where β_{it} are estimated in a first-pass as described above, and z_i contains different combinations of the variables: size, B/M, E/P, Asset/M, Asset/B

- Get final estimates as time-series averages of monthly estimates. Use standard error of the mean

- **Results in Table III**

β	ln(ME)	ln(BE/ME)	ln(A/ME)	ln(A/BE)	E/P Dummy	E(+)/P
0.15 (0.46)						
	−0.15 (−2.58)					
−0.37 (−1.21)	−0.17 (−3.41)					
		0.50 (5.71)				
			0.50 (5.69)	−0.57 (−5.34)		
					0.57 (2.28)	4.72 (4.57)
	−0.11 (−1.99)	0.35 (4.44)				
	−0.11 (−2.06)		0.35 (4.32)	−0.50 (−4.56)		
	−0.16 (−3.06)				0.06 (0.38)	2.99 (3.04)
	−0.13 (−2.47)	0.33 (4.46)			−0.14 (−0.90)	0.87 (1.23)
	−0.13 (−2.47)		0.32 (4.28)	−0.46 (−4.45)	−0.08 (−0.56)	1.15 (1.57)

- **Row 1:** weak explanatory power of beta (insignificant risk premium)
  - Note: on more recent data, the slope on beta is even *negative*

- There are **size and B/M effects**, even controlling for beta (rows 2, 3, and 4)
- **B/M prevails** on size effect (row 7)
- E/P effect *subsumed* by B/M
- Leverage: yes, with market leverage; no, with book leverage. The two coefficients are symmetric: it's really a B/M effect

$$c \log(A/M) - c \log(A/B) = c \log(B/M)$$

- The puzzling evidence that beta does not relate to average returns is picked up by Frazzini and Pedersen (2014)

Suggested Interpretations

1. **Rational Story:**
   - Size and B/M proxy for underlying risks which are associated with the behavior of earnings of small and value firms
   - Chan and Chen (1991, JF) show that small firms have worse performance under several profitability measures. Some of them are 'fallen angels'
   - F&F (1995, JF) show that there are size and value factors in earnings too
   - Argument for *'distress risk factor'* (ICAPM story)

2. **Behavioral Story:**

   - Correction of previous over-reaction (mean-reversion)
   - Idea: some firms indeed display negative earnings for a number of periods $\longrightarrow$ investors over-react and drive prices down too much because they incorrectly expect continuation of bad performance (B/M $\uparrow$). But then earnings mean-revert and investors are positively surprised: P $\uparrow$ and $R > 0$. Vice versa for growth stocks. This is the view of Lakonishok, Shleifer and Vishny (LSV, 1994, JF),
   - La Porta and LSV (1997, JF) confirm that value stocks have consistently positive earnings surprises and growth stocks – which they call glamour stocks – negative surprises. This is consistent with errors in expectations, rather than a risk premium

Fama and French, 1993, JFE

- Their previous paper shows that Size and B/M matter in explaining the cross-section of returns
- **Risk factors vs. characteristics?**
- If risk factors, need to show *covariation* of size and B/M portfolios with asset returns
  - Kozak, Nagel and Santosh (2018) for why this is not a sufficient condition for rational pricing
- To provide evidence of comovement of returns, need to use time series approach
- In time series approach, a test of the AP model is a test that the alphas are equal to zero
- Need the factors to be excess returns

The three factors

- One factor is the excess return on the market: $R_{m,t} - R_f$
- Two other factors come from double sorting of stocks along size and B/M dimensions:
  - 2 groups by size: Small, Big (NYSE breakpoints)
  - Independently, 3 groups by B/M: High, Medium, Low
  - Form 6 value-weighted portfolios from intersection

Size \ BM	Low	Medium	High
Big	LB	MB	HB
Small	LS	MS	HS

– Then:

$$\begin{aligned} HML &= \frac{1}{2}(HS + HB) - \frac{1}{2}(LS + LB) \\ SMB &= \frac{1}{3}(LS + MS + HS) \\ &\quad -\frac{1}{3}(LB + MB + HB) \end{aligned}$$

- HML ideally captures effect B/M controlling for size. SMB captures effect of size controlling for B/M

- **The three factors are excess returns**. Hence, you can use **time-series asset pricing tests**

The test assets

- Twenty-five value-weighted portfolios: 5×5 sorting by size and B/M
- Like before, B/M is measured in December of year $t-1$, size is measured in June of year t, and portfolios are formed annually in July of year t
- Sample: 1963-1991
- Summary statistics: average (monthly) returns increase with B/M and decrease with size

$$\begin{aligned} \bar{R}^e_{s=1,bm=5} &= 1.01\% \\ \bar{R}^e_{s=5,bm=1} &= 0.40\% \end{aligned}$$

A tautology?

- Possible doubt: explain B/M and size sorted portfolios using B/M and size sorted factors. Is it not a tautology?

- No! The factors only produce significant loadings *if there is genuine common variation* in the returns of the sorted portfolios

- If returns of size/B/M sorted portfolios covary, it must be because they share a common factor. With independent idiosyncratic risk and a large number of stocks, **idiosyncratic components wash out and only common factors drive portfolio covariation**

Results on covariation

- Look at loadings on factors and R^2 from time-series regressions

Table 4

Regressions of excess stock and bond returns (in percent) on the excess stock-market return, $RM-RF$: July 1963 to December 1991, 342 months.[a]

$$R(t) - RF(t) = a + b[RM(t) - RF(t)] + e(t)$$

Size quintile	Book-to-market equity (BE/ME) quintiles					Book-to-market equity (BE/ME) quintiles				
	Low	2	3	4	High	Low	2	3	4	High
	b					$t(b)$				
Small	1.40	1.26	1.14	1.06	1.08	26.33	28.12	27.01	25.03	23.01
2	1.42	1.25	1.12	1.02	1.13	35.76	35.56	33.12	33.14	29.04
3	1.36	1.15	1.04	0.96	1.08	42.98	42.52	37.50	35.81	31.16
4	1.24	1.14	1.03	0.98	1.10	51.67	55.12	46.96	37.00	32.76
Big	1.03	0.99	0.89	0.84	0.89	51.92	61.51	43.03	35.96	27.75
	R^2					$s(e)$				
Small	0.67	0.70	0.68	0.65	0.61	4.46	3.76	3.55	3.56	3.92
2	0.79	0.79	0.76	0.76	0.71	3.34	2.96	2.85	2.59	3.25
3	0.84	0.84	0.80	0.79	0.74	2.65	2.28	2.33	2.56	2.90
4	0.89	0.90	0.87	0.80	0.76	2.01	1.73	1.84	2.21	2.83
Big	0.89	0.92	0.84	0.80	0.69	1.66	1.35	1.73	1.95	2.69

[a] Dependent variable: Excess returns on 25 stock portfolios formed on size and book-to-market equity.

- Compare one-factor (CAPM, Table 4) and three-factor model (Table 6)

CAPM: $R^2 \approx 61\% - 92\%$. Value stocks have low betas compared to growth stocks, but higher average returns (value premium)

Size quintile	Book-to-market equity				
	Low	2	3	4	High
			b		
Small	1.04	1.02	0.95	0.91	0.96
2	1.11	1.06	1.00	0.97	1.09
3	1.12	1.02	0.98	0.97	1.09
4	1.07	1.08	1.04	1.05	1.18
Big	0.96	1.02	0.98	0.99	1.06
			s		
Small	1.46	1.26	1.19	1.17	1.23
2	1.00	0.98	0.88	0.73	0.89
3	0.76	0.65	0.60	0.48	0.66
4	0.37	0.33	0.29	0.24	0.41
Big	−0.17	−0.12	−0.23	−0.17	−0.05
			h		
Small	−0.29	0.08	0.26	0.40	0.62
2	−0.52	0.01	0.26	0.46	0.70
3	−0.38	−0.00	0.32	0.51	0.68
4	−0.42	0.04	0.30	0.56	0.74
Big	−0.46	0.00	0.21	0.57	0.76

- 3-factor model, factor loadings:
  - on the market $b \approx 1$ for all assets. It's what you'd expect, because the market factor captures orthogonal variation in returns of size and B/M sorted portfolios after controlling for component due to exposure to long-short (i.e., almost zero-beta) size and B/M factors

- loading on HML, $h > 0$ for value and $h < 0$ for growth
- loading on SMB, $s > 0$ for small and $s < 0$ for big

	R^2				
Size quintile	Low	2	3	4	High
Small	0.94	0.96	0.97	0.97	0.96
2	0.95	0.96	0.95	0.95	0.96
3	0.95	0.94	0.93	0.93	0.93
4	0.94	0.93	0.91	0.89	0.89
Big	0.94	0.92	0.88	0.90	0.83

- 3-factor model, R^2 increases substantially with three-factor model: **evidence of covariation of returns with two new factors**
- $R^2 \approx 83\% - 97\%$ — **almost complete spanning** (APT!?)

Results on AP performance

- The question is whether the model explains returns entirely ($\alpha_i = 0, \forall i$)
- Compare intercepts from CAPM (Table 9a, Panel (ii)) and the three-factor model (Table 9a, Panel (iv))

	Book-to-market equity (BE/ME) quintiles									
	a					$t(a)$				
Size quintile	Low	2	3	4	High	Low	2	3	4	High
(ii)	$R(t) - RF(t) = a + b[RM(t) - RF(t)] + e(t)$									
Small	−0.22	0.15	0.30	0.42	0.54	−0.90	0.73	1.54	2.19	2.53
2	−0.18	0.17	0.36	0.39	0.53	−1.00	1.05	2.35	2.79	3.01
3	−0.16	0.15	0.23	0.39	0.50	−1.12	1.25	1.82	3.20	3.19
4	−0.05	−0.14	0.12	0.35	0.57	−0.50	−1.50	1.20	2.91	3.71
Big	−0.04	−0.07	−0.07	0.20	0.21	−0.49	−0.95	−0.70	1.89	1.41
(iv)	$R(t) - RF(t) = a + b[RM(t) - RF(t)] + sSMB(t) + hHML(t) + e(t)$									
Small	−0.34	−0.12	−0.05	0.01	0.00	−3.16	−1.47	−0.73	0.22	0.14
2	−0.11	−0.01	0.08	0.03	0.02	−1.24	−0.20	1.04	0.51	0.34
3	−0.11	0.04	−0.04	0.05	0.05	−1.42	0.47	−0.47	0.71	0.56
4	0.09	−0.22	−0.08	0.03	0.13	1.07	−2.65	−0.99	0.33	1.24
Big	0.21	−0.05	−0.13	−0.05	−0.16	3.27	−0.67	−1.46	−0.69	−1.41

- **Size and B/M effects in CAPM:** significantly positive alphas for small and especially value portfolios. For example, $\alpha_{s=1,bm=5} = 0.54$ (t-stat=2.53)

- Instead, **with 3-factor model alphas of small and value portfolios are closer to zero**

  – Still, some problems with small-growth (1,1) portfolio. According to F&F these are less liquid stocks: no model would work

- **Test for joint significance** of the alphas (GRS statistic): Table 9c (column (ii) is CAPM, column (iv) is three-factor model)

	Regression (from tables 9a and 9b)				
	(i)	(ii)	(iii)	(iv)	(v)
F-statistic	2.09	1.91	1.78	1.56	1.66
Probability level					
Bootstrap	0.998	0.996	0.985	0.951	0.971
F-distribution	0.999	0.996	0.990	0.961	0.975

- The table shows $1 - Pvalue$. Hence, **all the asset pricing models are rejected**

- F&F don't make a big deal out of it. They argue that rejection is due to good fit of their model which reduces residual variance and increases power ($\Sigma \downarrow$ and $\alpha'\Sigma^{-1}\alpha \uparrow$: *rejection due to variance and not to level of* α)

What are size and B/M factors (rational stories)?

Distress risk factor story

- Preferred by F&F
- Stocks become value (high B/M) after a string of bad news; on the verge of distress, they are more exposed to *aggregate* shocks
- This must be **aggregate** (not idiosyncratic) distress: in Merton's ICAPM, they are poor hedges for deteriorating investment opportunities and must earn a risk premium orthogonal to market risk
- Consistent evidence: Liew and Vassalou (2000, JFE) show HML and SMB positively correlate with GDP growth, even controlling for the market

Background risk story

- Heaton and Lucas (2000, JF): average investor owns a small firm and rationally shies away from additional small/value risk (**background risk** story)

APT story

- Since the three-factor model explains almost 100% of return variation, by arbitrage arguments expected returns must be combinations of factor returns. The APT does *not* require the factors to be fundamental sources of risk

- Fama and French (1993) propose the covariance-with-risk-factor story as rational pricing. **Kozak et al. (2018) show this covariance logic does not pin down whether premia reflect risk or mispricing** (see later)

2. Daniel and Titman's Reply To F&F

The terms of the debate on the value premium

1. **Rational Story** (Fama and French):
   - Value companies load on a distress factor
   - The market requires high expected return because of high risk of this factor
   - Supporting evidence: **high covariance of firms with similar B/M**
2. **Behavioral Story** (Lakonishok, Shleifer, and Vishny, 1994, JF):
   - Investors extrapolate past performance into the future
   - Earnings mean-revert
   - Investors are surprised, which causes abnormal returns
   - It is possible that value firms covary–e.g., because they are in the same industry–but **risk premia are simply too high to be rationally justified by covariance with risk factor**
   - Problem with last statement: measuring appropriate risk premium requires full specification of a consumption based model

A different approach: Characteristics vs. Covariances

- Instead, Daniel and Titman (1997, JF) take a different line of attack against the rational view
- They want to show that the value premium is purely due to the **characteristic** of being high B/M and **not to the covariance** with any particular factor
- This approach is subject to the criticism of Kozak, Nagel, and Santosh (2018) discussed below
- **Approach:**
  - Identify firms with similar characteristics (B/M) but different loadings on risk factor (HML)
  - If covariance matters, they should have different average returns in spite of similar characteristic
  - **They don't!**

Methodology

- Construct 9 portfolios from a 3×3 sort by size and B/M
- Each of these 9 portfolios is further split into 5 on the basis of the expected $\beta_{i,hml}$, estimated on prior data
- Test: look at α_i from three-factor model for these 45 portfolios
  - **Risk-based model** predicts: $\alpha_i = 0$
  - **Characteristic-based model** (behavioral) predicts

$$\begin{aligned} \alpha_i &= E(R_i) - \\ &\quad -\beta_{i,hml}\lambda_{hml} - \beta_{i,smb}\lambda_{smb} - \beta_{i,mkt}\lambda_{mkt} \\ &= \begin{cases} > 0 \text{ for low } \beta_{i,hml} \\ < 0 \text{ for high } \beta_{i,hml} \end{cases} \end{aligned}$$

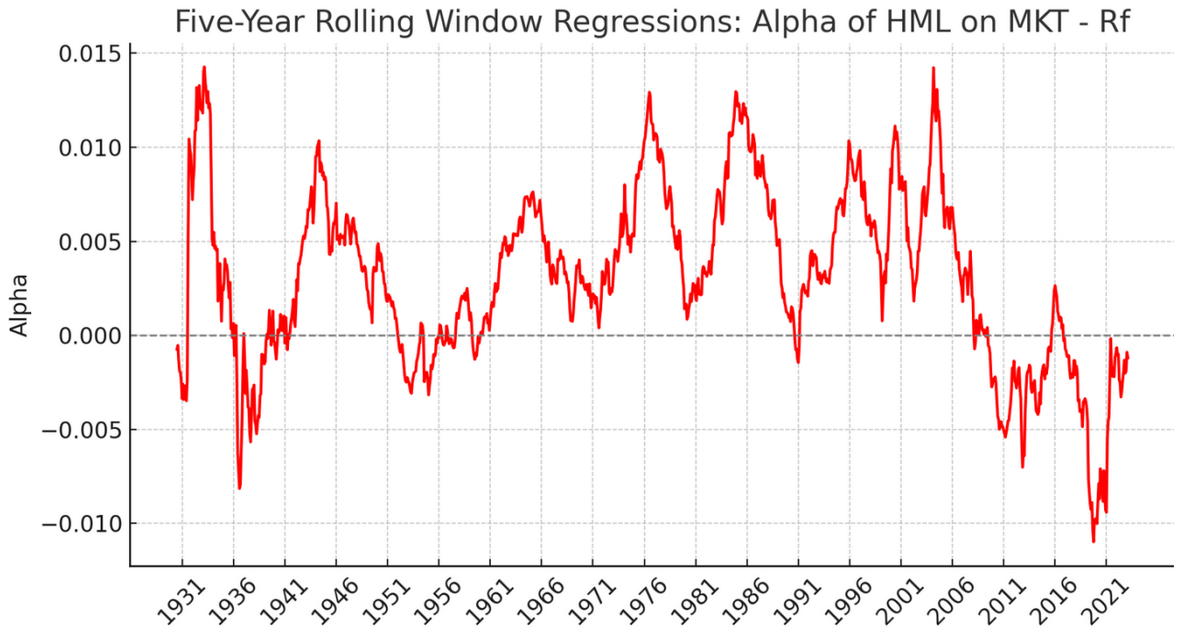
because λ_{hml}, λ_{smb}, and λ_{mkt} are positive and no relation is expected between $E(R_i)$ and the factor loadings

Char Port		Factor Loading Portfolio					Factor Loading Portfolio				
BM	SZ	1	2	3	4	5	1	2	3	4	5
				α					$t(\alpha)$		
1	1	−0.58	0.14	0.06	−0.17	−0.67	−3.97	1.04	0.48	−1.34	−4.00
1	2	0.16	0.05	0.13	0.16	−0.08	0.94	0.47	1.12	1.33	−0.63
1	3	0.02	0.05	0.06	−0.06	0.28	0.15	0.42	0.50	−0.55	2.26
2	1	0.13	0.08	0.06	0.21	−0.31	1.05	0.87	0.73	2.13	−2.31
2	2	0.03	0.20	0.22	0.01	−0.31	0.24	1.71	2.05	0.14	−2.66
2	3	0.19	−0.08	0.05	−0.10	−0.07	1.13	−0.50	0.35	−0.67	−0.46
3	1	0.08	0.05	0.10	0.01	−0.47	0.70	0.55	1.06	0.10	−3.27
3	2	0.17	0.22	0.25	0.05	−0.31	1.25	1.67	1.94	0.36	−1.63
3	3	−0.01	0.16	−0.23	−0.12	−0.18	−0.04	1.13	−1.45	−0.74	−0.90
Average		0.02	0.10	0.08	0.00	−0.24	0.16	0.82	0.75	0.08	−1.51

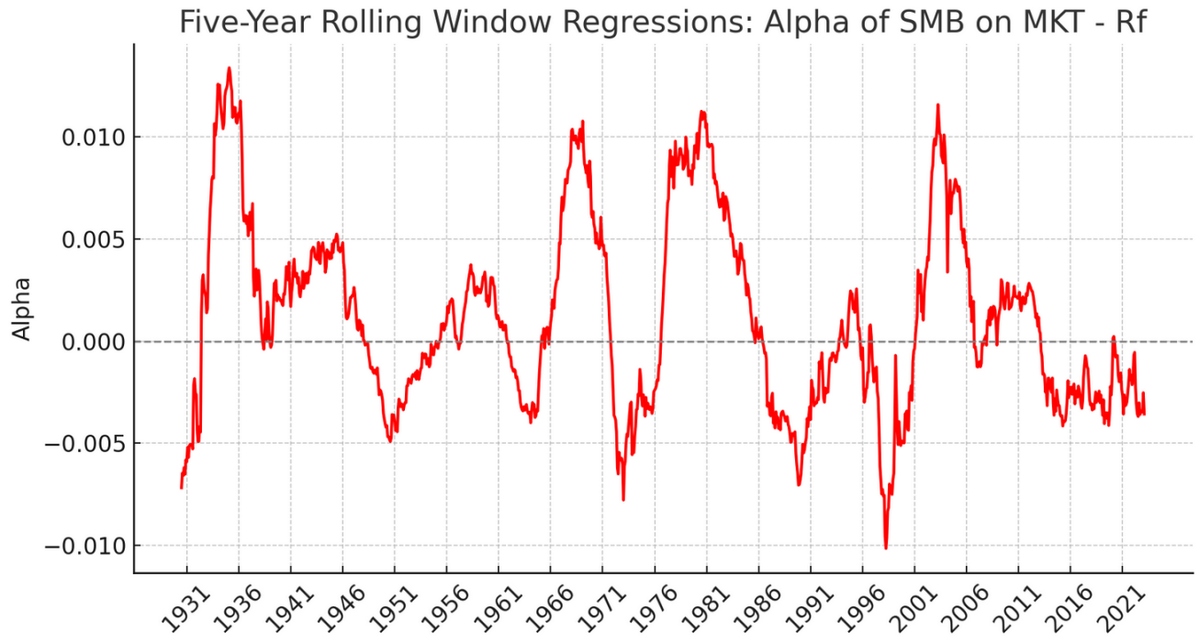
- **Results** in Table V (1973-1993):
  - Consistent with characteristics pricing, high $\beta_{i,hml}$ stocks have negative and significant α_i and vice versa
- **Conclusion:** *characteristics* (B/M) rather than covariance with risk factor ($\beta_{i,hml}$) explain expected returns
  - The fact that value and growth stocks move together is probably due that they are in the same sectors, which become in- and out-of-favor at a given point in time

More Recent Evidence on Size and B/M

- **Importantly, the value premium has disappeared starting after the Great Financial Crisis**
  - However, valuation-based mispricing has not disappeared. See Gonçalves and Leonard (2023) and Bergen, Franzoni, Obrycki, and Resendes (2025)



- SMB lost its significant alpha soon after it was discovered



3. Other Anomalies

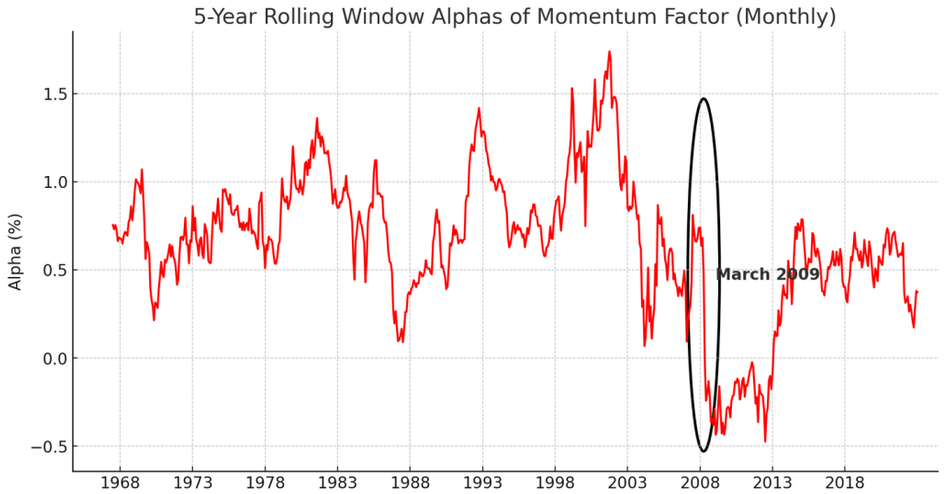
Reversal

- De Bondt and Thaler (1995, JF) form portfolios on the basis of past 3 to 5 year returns
- They show that **long term losers outperform long term winners** in the next 3 to 5 years
- A string of bad news gives a low price and a high B/M
- Hence, given the value premium, a stock with low past returns should earn high returns in the future
- **In the end, the reversal anomaly is really related to the B/M anomaly**

Momentum

- Jegadeesh and Titman (1993, JF) show that trading strategies based on short term losers and winners generate superior returns
- **Cross-sectional** anomaly
- Specifically, each month, buy stocks that are winners (in the past 3 to 12 months, skipping one month) and hold them for the next 3 to 12 months. Symmetrically, sell the short-term losers. **On average, the abnormal return is about 1% monthly** (1965-1989)
- This anomaly is **not explained** by the F&F three-factor model (F&F, 1996, JF), it survives after transaction costs (Korajczyk and Sadka, 2004 JF), and it is present in later samples (Jegadeesh and Titman, 2001 JF)
- **Momentum is among the most persistent anomalies**
- The literature has evolved to include a momentum factor alongside the Fama-French factors, e.g., in mutual fund performance evaluation (Carhart, 1997, JF)

- **There are momentum crashes** following the end of a down market, e.g. 2009
  - During a bear market the portfolio is long the low beta stocks (winners). When the market turns around the low beta stocks become losers and the portfolio experiences a drawdown



Possible (behavioral) stories

1. **Overreaction**
   - *Positive-feedback traders* invest in winners and sell losers driving pricing away from fundamentals for some time
   - Long-run reversal would be evidence in favor of this story
2. **Underreaction**
   - Investor do not fully incorporate information about short-term prospects of the firms, such as information in earnings announcement
   - Consistent with this story, Chan, Jegadeesh, and Lakonishok (1996, JF) find **earnings momentum**: one can predict future returns based on past earnings surprises. This is also consistent with the **post-earnings-announcement drift** (Bernard and Thomas, 1989, J. of Accounting and Economics)

- The theoretical paper by Hong and Stein (1999, JF) keeps the two stories together by postulating slow information diffusion and two types of bounded rational traders:
  - **News-watchers**: cause underreaction because they don't extract other traders' information
  - **Trend-chasers**: cause overreaction because they condition their trades only on past price movements

- Behavioral stories need **limits to arbitrage** to survive. Moskowitz and Grinblatt (1999, JF) find that momentum is mainly an industry phenomenon (winner and loser industries). Hence, momentum returns contain industry factors which are hard to diversify away
  - No perfect arbitrage, consistent with Kozak, Nagel, and Santosh (2018). See later

Idiosyncratic Volatility is Priced

(Ang, Hodrick, Xing, and Zhang, 2006 JF, 2009 JFE)

- AHXZ (2006, JF): **low-ivol stocks outperform high-ivol stocks** in the US by 1% per month, unexplained by F&F factors, momentum, or VIX sensitivity
- AHXZ (2009, JFE) extend the analysis to 23 countries to rule out data mining and test for international comovement
- **Ivol construction:** each month t, regress daily returns of stock i on F&F factors using data from month $t-1$:

$$r_{i,t} = \alpha_i + \beta_i MKT_t + s_i SMB_t + h_i HML_t + \varepsilon_{i,t}$$

  ivol $=$ **std. dev. of residuals** $\hat{\varepsilon}_i$; factor loadings used as risk controls

- **Methodology:** Fama–MacBeth regressions of returns on ivol, factor loadings, and controls (size, B/M, past 6-month return, country dummies):

$$r_{i,t} \;=\; c + \gamma\, ivol_{i,t} + \lambda'_\beta \beta_{i,t} + \lambda'_z z_{i,t} + u_{i,t}$$

Results

- **Cross-sectional evidence (Table 2):**

	Canada	France	Germany	Italy	Japan	U.K.	U.S.
Panel A: USD Denominated Returns							
Constant	1.723	0.602	0.753	0.425	0.948	0.480	1.746
	[3.68]	[1.13]	[1.87]	[0.76]	[1.25]	[1.03]	[3.83]
W-FF Idiosyncratic Volatility	-1.224	-1.439	-2.003	-1.572	-1.955	-0.871	-2.014
	[-2.46]	[-2.14]	[-3.85]	[-2.10]	[-5.18]	[-2.54]	[-6.67]
$\beta(MKT^W)$	0.344	0.059	0.277	-0.083	0.323	0.178	0.376
	[2.20]	[0.44]	[1.93]	[-0.32]	[3.12]	[1.46]	[4.52]
$\beta(SMB^W)$	0.009	0.015	-0.083	0.116	0.050	0.032	-0.049
	[0.12]	[0.17]	[-0.82]	[0.56]	[0.76]	[0.42]	[-1.19]
$\beta(HML^W)$	-0.070	-0.069	0.076	-0.221	-0.025	-0.077	-0.051
	[-0.95]	[-0.94]	[1.00]	[-1.98]	[-0.35]	[-1.30]	[-1.69]
Size	-0.253	-0.067	-0.044	-0.031	-0.132	-0.058	-0.157
	[-4.81]	[-1.08]	[-1.09]	[-0.47]	[-1.72]	[-1.16]	[-3.14]
Book-to-Market	0.369	0.569	0.176	0.239	0.550	0.365	0.282
	[3.68]	[4.59]	[1.35]	[1.48]	[3.84]	[4.46]	[3.87]
Lagged Return	0.014	0.001	0.003	0.001	-0.011	0.012	-0.001
	[3.57]	[0.10]	[1.01]	[0.15]	[-2.85]	[4.07]	[0.28]
Adjusted R^2	0.118	0.108	0.114	0.147	0.124	0.078	0.046
Percentiles of W-FF Idiosyncratic Volatility							
25th Percentile	20.8	21.4	16.3	21.5	23.1	13.9	25.0
75th Percentile	46.0	39.2	34.8	38.4	39.6	31.3	61.1
Economic Effect of Moving from the 25th to the 75th W-FF Idiosyncratic Volatility Percentiles							
25% → 75%	-0.31%	-0.26%	-0.37%	-0.27%	-0.32%	-0.15%	-0.73%

  - **Ivol is negative and significant for all countries**

  - Magnitude: strongest in US — moving from 25th to 75th percentile of ivol causes a **0.73% drop** in monthly returns

- **Portfolio evidence:** within-country ivol-quintile portfolios aggregated at the regional level yield negative and significant 3-factor alphas in all regions (strongest in US: -1.95%/month)

- A US ivol factor (VOL^{US}) captures international comovement and renders regional alphas insignificant — suggestive of a common risk factor, but the authors stop short of that interpretation without an economic motivation

An explanation of the IVOL anomaly (Stambaugh, Yu, and Yuan, JF 2015)

- SYY argue the anomaly persists due to **two frictions**:
  - *Arbitrage risk*: high-IVOL stocks are harder to arbitrage, so mispricing – whether overpricing or underpricing – is more persistent for them
  - *Arbitrage asymmetry*: far more capital (e.g. long-only mutual funds) corrects underpricing than overpricing, so overpricing is larger in magnitude than underpricing for high-IVOL stocks
- Net effect: high-IVOL stocks are on average overpriced $\Rightarrow$ low subsequent returns. Mispricing is exogenous (sentiment traders); IVOL acts as a *limit to arbitrage*
- **Supporting evidence:** double sort on IVOL and a composite mispricing score (average rank across 11 anomalies: momentum, profitability, investment, accruals, etc.)

	Highest IVOL	Next 20%	Next 20%	Next 20%	Lowest IVOL	Highest –Lowest	All Stocks
Most Overpriced	−1.89	−0.95	−0.72	−0.47	−0.39	−1.50	−0.81
(Top 20%)	(−12.05)	(−7.39)	(−4.90)	(−3.62)	(−3.04)	(−7.36)	(−8.14)
Next 20%	−0.88	−0.41	−0.31	−0.21	−0.04	−0.84	−0.23
	(−5.86)	(−3.36)	(−3.00)	(−2.08)	(−0.44)	(−4.41)	(−3.88)
Next 20%	−0.09	−0.01	−0.05	−0.12	0.02	−0.10	−0.07
	(−0.53)	(−0.09)	(−0.48)	(−1.29)	(0.18)	(−0.53)	(−1.47)
Next 20%	−0.15	0.07	0.17	0.18	0.23	−0.38	0.18
	(−0.80)	(0.63)	(1.87)	(2.33)	(3.22)	(−1.78)	(4.45)
Most Underpriced	0.56	0.68	0.51	0.33	0.14	0.41	0.28
(Bottom 20%)	(3.27)	(4.91)	(5.02)	(4.10)	(2.04)	(2.16)	(5.67)
Most Overpriced –	−.44	−1.63	−1.23	−0.81	−0.53	−1.91	−1.09
Most Underpriced	(−11.07)	(−8.65)	(−6.43)	(−5.02)	(−3.43)	(−7.62)	(−8.05)
All Stocks	−0.69	−0.12	−0.00	0.05	0.08	−0.78	
	(−6.09)	(−1.56)	(−0.01)	(1.07)	(1.86)	(−5.50)	

- **Red box:** more overpricing (more negative α) for high-IVOL stocks – consistent with arbitrage risk
- **Blue box:** more underpricing (more positive α) for high-IVOL stocks – also consistent with arbitrage risk
- **Asymmetry:** $|-1.50| \gg 0.41$ – the overpricing dominates, consistent with arbitrage asymmetry
- **Green box:** aggregating across all stocks, overpricing prevails $\Rightarrow$ decreasing relation between returns and IVOL

Betting against Beta, Frazzini and Pedersen, 2014

Motivation and model intuition

- **Fact:** beta does not explain the cross-section of average returns (flat empirical SML)
- **Why?** CAPM assumes investors can freely borrow, but many cannot (e.g. mutual funds). Black (1972) had an early version of CAPM with borrowing constraints
- **Mechanism:** leverage-constrained investors reach for high expected returns by overweighting high-beta stocks $\Rightarrow$ high-beta prices are bid up and expected returns fall; low-beta stocks are neglected and earn a premium
- Agents that *can* leverage (e.g. hedge funds) partially arbitrage this away by buying low-beta and shorting high-beta, but margin constraints limit their ability to restore the correct SML
- **Result:** the SML is too flat – higher intercept than r_f and lower slope than $E(r_m) - r_f$ – so **low-beta stocks earn higher risk-adjusted returns than high-beta stocks**
- **Paper's contribution:** formalize this in a model; construct a **BAB factor** (long low-beta, short high-beta) that earns positive risk-adjusted returns across asset classes

The BAB factor: construction and empirical evidence

- **Construction** (zero-beta, zero-net-investment): pre-ranking betas from daily returns (one year, Dimson 1979 adjustment); every month form rank-weighted low- and high-beta portfolios r^L, r^H:

$$r_{t+1}^{BAB} = \frac{1}{\beta_t^L}\left(r_{t+1}^L - r_f\right) - \frac{1}{\beta_t^H}\left(r_{t+1}^H - r_f\right) \tag{1}$$

with $\beta^L = 0.7$ and $\beta^H = 1.5$ for US equity; long $\$\frac{1}{0.7}$ in low-beta, short $\$\frac{1}{1.5}$ in high-beta stocks

	P1 (Low beta)	P2	P3	P4	P5	P6	P7	P8	P9	P10 (high beta)	BAB Factor
Excess return	**0.99** (5.90)	**0.90** (5.24)	**0.92** (4.88)	**0.98** (4.76)	**1.04** (4.56)	**1.12** (4.52)	**1.07** (4.08)	**1.07** (3.71)	**1.03** (3.32)	**1.02** (2.77)	**0.71** (6.76)
CAPM alpha	**0.54** (5.22)	**0.39** (4.70)	**0.35** (4.23)	**0.35** (4.00)	**0.34** (3.55)	**0.37** (3.41)	**0.26** (2.45)	0.19 (1.54)	0.09 (0.65)	-0.05 -(0.29)	**0.69** (6.55)
3-factor alpha	**0.38** (5.24)	**0.25** (4.43)	**0.19** (3.69)	**0.18** (3.62)	**0.15** (2.65)	**0.14** (2.49)	0.04 (0.75)	-0.07 -(1.06)	**-0.18** -(2.45)	**-0.36** -(3.10)	**0.66** (6.28)
4-factor alpha	**0.42** (5.66)	**0.32** (5.67)	**0.24** (4.55)	**0.24** (4.63)	**0.24** (4.20)	**0.25** (4.58)	**0.17** (3.00)	**0.12** (1.98)	0.04 (0.61)	-0.07 -(0.59)	**0.55** (5.12)
5-factor alpha*	**0.23** (2.37)	**0.23** (3.00)	**0.17** (2.28)	**0.16** (2.13)	**0.16** (2.08)	**0.20** (2.76)	**0.22** (2.86)	0.06 (0.69)	0.11 (1.08)	0.01 (0.07)	**0.46** (2.93)
Beta (ex ante)	0.57	0.75	0.84	0.92	0.99	1.06	1.14	1.23	1.36	1.64	0.00
Beta (realized)	0.75	0.86	0.97	1.07	1.18	1.28	1.37	1.50	1.60	1.82	0.03
Volatility	18.2	18.7	20.6	22.4	24.7	27.0	28.4	31.5	33.8	40.0	11.5
Sharpe ratio	0.65	0.58	0.54	0.52	0.50	0.50	0.45	0.41	0.37	0.31	0.75

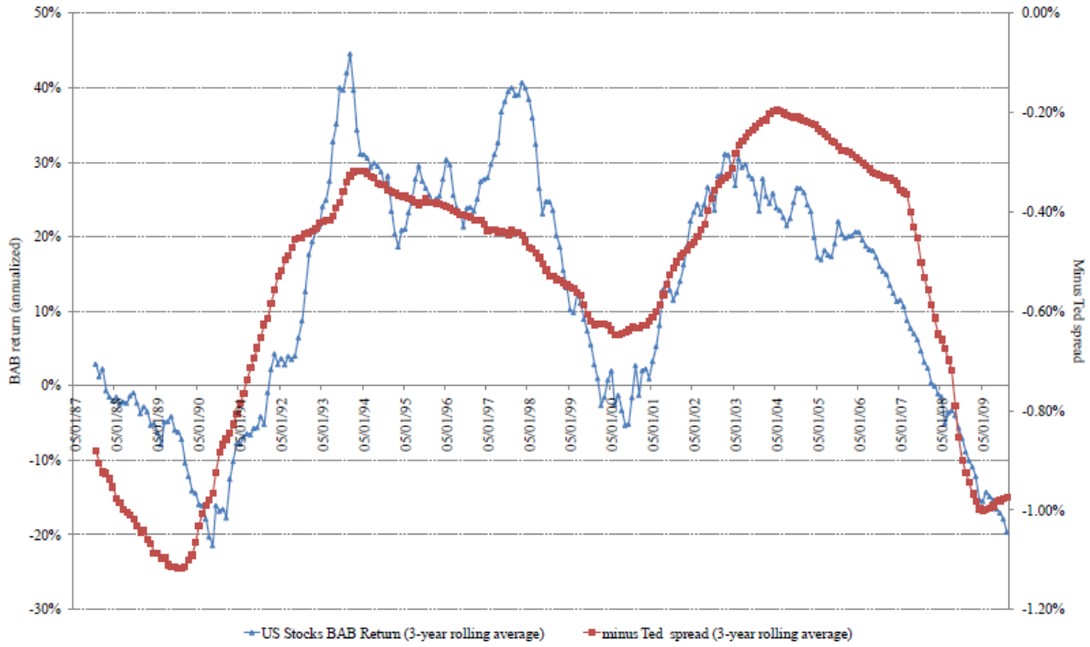
- Average excess returns are similar across the ten beta-sorted portfolios $\Rightarrow$ leveraging up low-beta and leveraging down high-beta yields a positive spread at near-zero market exposure

- **BAB has significant alpha** against the market, FF three-factor, Carhart four-factor, and Pastor–Stambaugh (with a liquidity factor) five-factor models

- Annual return: **11.5%**; annualized Sharpe ratio: **0.75** (comparable to momentum)

Further evidence: funding liquidity and portfolio holdings

- **Funding liquidity:** the TED spread (3m LIBOR − 3m T-bill) measures the availability of trading capital to leveraged speculators (Brunnermeier and Pedersen, 2009; Garleanu and Pedersen, 2011). When funding tightens, hedge funds are forced to sell low-beta positions, so BAB should have negative realizations – confirmed in the chart (which plots minus the TED spread)



- **Portfolio holdings:** consistent with the theory, leverage-constrained investors (mutual funds, individuals) overweight high-beta stocks, while investors that can borrow (LBO firms, Berkshire Hathaway) overweight low-beta stocks – both confirmed in the data

Critical assessment

- Given that the SML does not work, the BAB factor in equation (1) is built to have a positive alpha:
  - All portfolios have the same $E(r_i)$
  - Then, if you divide by $\beta_L < 1$ you get something large from which you subtract something small (because it is divided by $\beta_H > 1$)
  - This is simply Black's (1972) idea to exploit the fact that the SML is too flat
  - In this sense, the idea is not original
- "Betting Against Betting Against Beta", Novy-Marx and Velikov (2022, JFE)
  - The rank-weighted portfolios **overweight nano-caps** (those in the first decile of the size distribution)
  - Value-weighted monthly performance drops from 1% to 0.56%
  - The problem with overweighting small stocks is transaction costs: 60 bps/month (computed using Novy-Marx and Velikov, 2016, RFS)

- Net-of-transaction-costs monthly performance is 0.48%/month
- The strategy **loads significantly on the investment and profitability factors** (see below)
- The net-of-transaction-costs alpha of BAB from five-factor model is 0.06%/month

Fama and French: the five-factor model (JFE 2015, RFS 2016)

- **Motivating results**:
  - Novy-Marx's (2013): **profitability correlates positively with average returns**, controlling for other characteristics
  - Titman, Wei, and Xie's (2004): **investment** – i.e., assets growth – **correlates negatively with average returns**, controlling for other characteristics
- F&F add **two new factors** to the three-factor model:
  - **RMW** (robust minus weak): long high-profitability, short low-profitability stocks
  - **CMA** (conservative minus aggressive): long low-investment, short high-investment stocks

$$\begin{aligned} R_{i,t} &= a_i + b_i\left(R_t^{mkt} - R_t^f\right) + h_i HML_t \\ &\quad + s_i SMB_t + r_i RMW_t + c_i CMA_t + e_{i,t} \end{aligned}$$

- F&F's **Explanation– the DCF identity:**

- Start from the dividend discount model for the price of a stock

$$M_t = \sum_{\tau=1}^{\infty} \frac{E\left(D_{t+\tau}\right)}{(1+r)^{\tau}}$$

- Express dividends as a function of earnings and the change in book values: $D_{t+\tau} = Y_{t+\tau} - dB_{t+\tau}$
- Divide by book

$$\frac{M_t}{B_t} = \sum_{\tau=1}^{\infty} \frac{E\left(Y_{t+\tau}/B_t - dB_{t+\tau}/B_t\right)}{(1+r)^{\tau}} \tag{2}$$

  which implies (ceteris paribus): higher B/M $\Rightarrow$ higher r; higher profitability $\Rightarrow$ higher r; higher investment $\Rightarrow$ lower r. This is a pure identity – **no claim about risk vs. mispricing**

- **Results:** the model absorbs the low-beta (Frazzini–Pedersen), share-repurchase (Loughran–Ritter), and low-volatility (Ang et al.) anomalies via RMW and CMA loadings; **once RMW and CMA are included, HML loses all incremental explanatory power**

- **Limits:** the model cannot explain Accruals (Sloan, 1996) or Momentum (Jegadeesh–Titman, 1993)

A critique: the investment–profitability interaction (Franzoni, Obrycki, and Resendes, 2025)

- The DCF identity in Equation (2) treats investment as always reducing expected returns, but ignores the case of positive-NPV projects. A residual income reformulation makes this clear:

$$M_t = B_t + \sum_{i=1}^{\infty} \frac{(ROE_{t+i} - r)\, B_{t+i-1}}{(1+r)^i} \tag{3}$$

  Dividing by B_t and substituting $B_{t+i-1}/B_t = 1 + I_{t+i-1}$:

$$\frac{M_t}{B_t} = 1 + \sum_{i=1}^{\infty} \frac{(ROE_{t+i} - r)\,(1 + I_{t+i-1})}{(1+r)^i}$$

  When $ROE > r$, higher investment *increases* firm value $\Rightarrow$ investment is **positively** related to expected returns for high-profitability firms

- F&F's model misses this **Investment**$\times$**Profitability interaction**: a factor long high-profitability/high-investment and low-profitability/low-investment firms, and short the mixed pairs, earns five-factor alphas of **33 bps/month**

	Dep. Variable: Stock Return					
	All stocks		Large Stocks		Small Stocks	
	(1)	(2)	(3)	(4)	(5)	(6)
ROE	5.724	6.744	2.989	2.858	6.670	7.164
	(7.327)	(11.745)	(3.214)	(3.400)	(8.591)	(11.776)
Investment	-0.689	-0.498	-0.483	-0.447	-0.681	-0.497
	(-5.875)	(-5.697)	(-2.879)	(-3.014)	(-5.688)	(-5.356)
ROE×Inv.	4.637	4.097	6.087	5.190	3.874	4.021
	(2.920)	(3.479)	(1.970)	(1.925)	(2.727)	(3.367)
Beta		0.111		0.071		0.134
		(1.015)		(0.468)		(1.321)
Ret_1		-0.041		-0.027		-0.043
		(-11.617)		(-5.526)		(-12.075)
Ret_{2_13}		0.005		0.004		0.005
		(4.806)		(2.518)		(5.219)
log(BM)		0.250		0.091		0.273
		(4.578)		(1.403)		(4.608)
log(M)		-0.093		-0.035		-0.134
		(-2.830)		(-1.043)		(-2.762)
Accruals		-0.862		-0.466		-0.884
		(-2.745)		(-1.017)		(-2.539)
Constant	1.154	3.835	1.032	1.989	1.227	4.409
	(5.038)	(7.514)	(5.031)	(3.052)	(5.047)	(7.263)
N	1,388,625	1,388,625	352,353	352,353	1,036,272	1,036,272
R-squared	0.014	0.051	0.023	0.114	0.013	0.046

4. Kozak, Nagel, and Santosh, 2018: Interpreting Factor Models

Motivation

- There was a consensus that:

$$E(R) = \lambda'\beta \Longrightarrow \text{Rational Investors}$$

  where β are **covariances** of returns with a set of factors

- These authors argue that **tests of "charcteristics" vs. "covariances"** – e.g. Daniel and Titman, 1997 – **cannot discriminate between rational and behavioral theories of prices**

- More broadly, *any* test that tries to discriminate rational from behavioral pricing using return data alone is subject to this critique – including investment-based models (e.g. Lin and Zhang, 2013) and ICAPM-style tests (e.g. Campbell and Vuolteenaho, 2004)

- The paper is very sophisticated and worth studying in depth if you want to work in asset pricing

- Here, we try to summarize their logic

Absence of Near-Arbitrage Opportunities

- The Law of One Price (LOP) implies that there exists an SDF M that prices all assets

$$E(MR) = 0 \tag{4}$$

where R is an $N \times 1$ vector of returns in excess of the risk free rate and $\mu = E(R)$, $\Gamma = Var(R)$

- From Hansen and Jagannathan (1991), we know that the squared maximum Sharpe Ratio that one can achieve from a given set of assets is equal to the variance of the SDF (see proof on next slide)

$$Max\ SR^2 = Var(M)$$

- **Absence of Near-Arbitrage Opportunities means that the maximum Sharpe Ratio is not implausibly large** (e.g., above 2) and, therefore, that the SDF is not excessively volatile

Proof: $Max\ SR^2 = Var(M)$

- Among all the possible SDF, consider the SDF represented as a linear projection on the space of excess returns

$$M = 1 + b(R_x - E(R_x)) \tag{5}$$

where R_x is a mean-variance efficient portfolio, which always exists.
Note that because we are working with excess returns the constant a in the SDF is indetermined and we can choose so that $E(M) = 1$

- Re-write the pricing equation (5) using this definition of the SDF and apply it to price R_x itself

$$\begin{aligned} E((1 + b(R_x - E(R_x)))R_x) &= 0 \\ E(R_x) + bE\left(R_x^2\right) - bE^2(R_x) &= 0 \\ E(R_x) + bVar(R_x) &= 0 \end{aligned}$$

- From which we obtain

$$b = -\frac{E(R_x)}{Var(R_x)}$$

- Thus, we can compute $Var\left(M\right)$ as

$$\begin{aligned} Var\left(M\right) &= Var\left(1+b\left(R_x - E\left(R_x\right)\right)\right) \\ &= b^2 Var\left(R_x\right) \\ &= \frac{E^2\left(R_x\right)}{Var\left(R_x\right)} \\ &= SR_{R_x}^2 \end{aligned}$$

  which completes the proof.

Implications for the number of factors

- Express the var-cov of the returns Γ as a function of N principal components (PCs) via eigenvector decomposition

$$\Gamma = Q \Lambda Q'$$

with $Q = (q_1, q_2, ..., q_N)$

- **You can interpret the number of PC's that explain most of the variance of returns as the number of factors that explain returns**

- Kozak et al. show that the variance of the SDF is

$$Var\left(M\right) = \frac{\mu_m^2}{\sigma_m^2} + N \cdot Var_{cs}\left(\mu_i\right) \sum_{k=2}^{N} \frac{Corr_{cs}\left(\mu_i, q_{ki}\right)^2}{\lambda_k}$$

where $\frac{\mu_m^2}{\sigma_m^2}$ can be interpreted as the squared Sharpe Ratio of the market factor (i.e., the first eigenvector), while $Var_{cs}\left(\mu_i\right)$ and $Corr_{cs}\left(\mu_i, q_{ki}\right)$ denote cross-sectional moments, and λ_k denote eigenvalues associated with eigenvectors q_k more lowly ranked than the first eigenvector

  - The elements of q_k can be interpreted as portfolio weights

- **So, the $Var(M)$ and, therefore, the maximum squared Sharpe Ratio, is not too large only if the number of PCs that explain returns is not too large**

- To see that, interpret $Corr_{cs}(\mu_i, q_{ki})^2$ as the R-squared of a regression of average stock returns on a given PC. In turn, a PC captures an asset pricing anomaly – i.e., a cross-sectional dispersion in returns not explained by exposure to the market factor (the first PC). If $Corr_{cs}(\mu_i, q_{ki})^2$ is large for PCs that have low λ_k, i.e. more lowly ranked PCs, then the summation is large and the $Var(M)$ is large, leading to Near-Arbitrage Opportunities

- **Example:** suppose PC k resembles the HML factor – it has large positive weights on value stocks (i = high B/M) and large negative weights on growth stocks (i = low B/M). Then $Corr_{cs}(\mu_i, q_{ki})$ is large and positive if value stocks have high expected returns and growth stocks have low expected returns – i.e., the value premium exists. $Corr_{cs}(\mu_i, q_{ki})^2$ is therefore the share of the cross-sectional dispersion in μ_i explained by that PC: it equals 1 if the value premium is the *only* source of cross-sectional return variation, and 0 if HML is entirely unpriced

- **Intuition:** if expected returns were concentrated in *low-variance* directions of the return covariance matrix – i.e., low λ_k– one could form a high-Sharpe portfolio at very low risk – approaching arbitrage.

- **Implication**: Absence of near-arbitrage therefore *forces* expected returns to align with the high-variance principal components, regardless of whether those returns originate from rational risk premia or investor sentiment

- **Conclusion: Absence of Near Arbitrage corresponds to a few factors explaining returns**

What implications for the Economic Model?

- **This argument does not tell us whether asset prices are determined by rational or behavioral investors**
- The authors build a model where sentiment driven investors trade with rational arbitrageurs
- They impose Absence of Near Arbitrage Opportunities, corresponding to limits on the amount of mispricing that sentiment-driven investors can cause
- **They show that the resulting SDF contains components related to sentiment-driven demand**
- Importantly, only if the sentiment-driven components correlate with the most important PCs, do they have a significant impact on the SDF and, therefore, they are factors of pricing concern
- **Intuition:** If sentiment investors' demands do not covary with the most important sources of risk – i.e. the first few PCs – arbitrageurs can trade aggressively against them, without taking on significant risk, and cause the deviations from CAPM pricing to disappear. **The sentiment demand that has pricing impact is only the one that covaries with important PCs**

- **Example 1 (no pricing impact):** sentiment investors are irrationally bullish on a narrow cluster of biotech stocks with low market covariance. Arbitrageurs short them cheaply – little factor risk – and prices revert. The sentiment demand is orthogonal to the dominant PCs and gets fully arbitraged away

- **Example 2 (pricing impact):** sentiment investors systematically overvalue growth relative to value stocks – a demand pattern aligned with HML, a dominant PC. Arbitrageurs who trade against this must go long value and short growth, exposing themselves to large factor swings. The risk deters full arbitrage, the value premium persists, and it *looks* like a rational risk premium even though it originates from sentiment

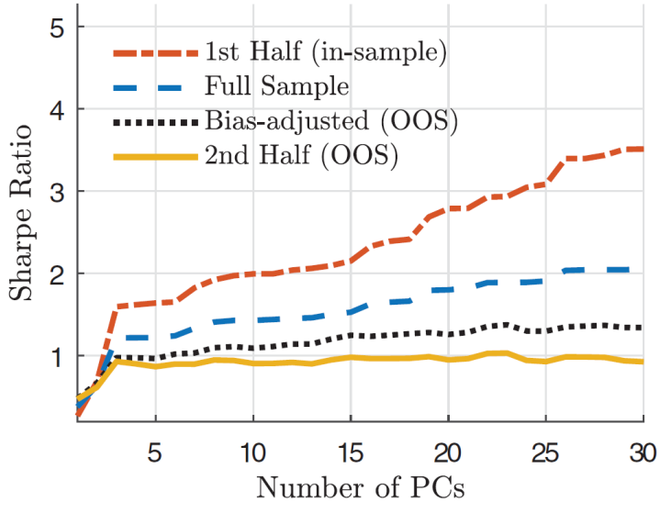
- This argument is related to the limits of arbitrage literature (see later): if arbitrageurs' wealth covaries with sentiment, arbitrageurs cannot trade against the sentiment

- **Bottom line: Covariances of anomaly returns with factors do not necessarily imply absence of sentiment driven investors who affect prices**

- **Consequence for the characteristics-vs-covariances debate:** in this equilibrium, expected returns still align with factor covariances – even though pricing is driven by sentiment. Characteristics and covariances therefore carry *identical* information about expected returns. **A horse race between them has no power to detect the presence of sentiment investors**

How do they explain prior evidence?

- How is it possible that more and more characteristics have appeared in previous papers that are not explained by a few factors?
- They argue that evidence of characteristics being priced which are not explained by few PCs **does not survive out-of-sample** (OOS)
- It could be the result of data mining or of temporary mispricing due to arbitrageurs' **learning** about the anomaly
- They show consistent evidence in Figure 5 where they plot the in-sample and out-of-sample maximum Sharpe ratios (annualized) obtained combining the first K principal components of 30 anomaly (long and short) portfolio returns
  - The in-sample Sharpe Ratio becomes very large by adding PCs – i.e., evidence that are not explained by just a few PCs $\longrightarrow$ Consistent with arbitrage opportunities
  - However, the out-of-sample Sharpe Ratio does not increase beyond the first 3 PCs



Panel A: 30 anomaly portfolios (sample split)

- **Connection to McLean and Pontiff (2016):** their finding that anomaly returns decay by 58% after publication is consistent with the KNS framework – the higher-order PCs that appear priced in sample reflect either data mining or mispricing that arbitrageurs gradually eliminate once the anomaly becomes public knowledge. The out-of-sample collapse of the Sharpe ratio in Figure 5 is the KNS counterpart to that post-publication decay. See EMH notes

5. Replication Crisis

Overview

- There is a proliferation of asset pricing anomalies in the first decade and a half of the 2000's, leading Cochrane (2011) to talk about the **factor zoo**
- Spurred by this proliferation of factors, some researchers have questioned the possibility of finding so many predictors, e.g. such a large degree of market inefficiency
- This literature has been inspired by the new trend towards replicating past empirical results that follows the Replication Crisis in Psychology
- Different authors have followed different approaches to **tame the factor zoo**:
  - **Out-of-sample testing**. We have already discussed McLean and Pontiff (2016, JF). A similar paper : Linnainmaa and Roberts (2018) study **pre-sample** replicability of several accounting-based anomalies and reach the conclusion that few anomalies survive
  - **Adjustments** for the tradability of the strategy. An example is Hou, Xue, and Zhang (2020)
  - **Statistical correction** for multiple testing. Harvey, Liu, and Zhu (2016, RFS) vs. Jensen, Kelly, and Pedersen (2023, JF)

– Another literature, instead, has applied **Machine Learning** to Asset Pricing to let the data speak on the variables that predict returns and try to reduce the dimensionality of the number of predictors

Hou, Xue, and Zhang (2020, RFS): The Microcap Problem

- They study replicability of **452 anomalies**, arguing that many results are driven by **microcaps** – stocks smaller than the 20th percentile of the market equity distribution for NYSE stocks – which are **not really tradable**
  - Microcaps are only 3.2% of aggregate market capitalization but 60.7% of the number of stocks; they have the highest equal-weighted returns and the largest cross-sectional dispersions in returns and anomaly exposure
  - Many original studies **overweight** microcaps via **equal-weighted** returns and **NYSE-Amex-NASDAQ breakpoints**; hundreds also use **OLS cross-sectional regressions**, which are highly sensitive to microcap outliers
- **Their corrections:** use **NYSE breakpoints** and **value-weighted** returns in portfolio analyses; use **weighted least squares** (market-cap weights) in cross-sectional regressions

Hou, Xue, and Zhang (2020, RFS): Results and Lessons

- **Most anomalies fail to replicate:** after controlling for microcaps, **65%** of the 452 anomalies cannot clear the single-test hurdle of $|t| \geq 1.96$; imposing the multiple-testing hurdle of 2.78 (Harvey, Liu, and Zhu 2016, discussed below) raises the failure rate to **82%**

- Even for anomalies that do replicate, their **economic magnitudes are much smaller** than originally reported

- **Lessons:** the study highlights the roles of **publication bias** (only significant results get published, with no incentive to replicate) and **p-hacking** (strong incentives to search over many specifications until significance is found) – both of which call for **raising the threshold for statistical significance**

Harvey, Liu, and Zhu (2016, RFS)

- They advocate a statistical correction for **multiple-hypotheses testing**
- Point of view: Even though each individual researcher carries out his/her own tests in isolation, collectively the finance research profession is acting as if it was running **multiple tests on the same data**
- **Intuition:** Suppose you flip a fair coin 20 times. The probability of getting 14+ heads by pure chance is about 5%. Now imagine 100 researchers each doing this independently and only publishing when they get 14+ heads. Even though every coin is fair, **roughly 5 papers will claim the coin is biased** – purely by chance. The same logic applies when many researchers independently test return predictors on the same market data
- In other disciplines, e.g. physics or medicine, multiple testing corrections are standard
  - E.g., in a lab experiment that collects hundreds of signals, the researcher corrects for the fact that the more signals are tested the more are going to result as statistically different from zero

- Hence, the authors prescribe **higher thresholds of statistical significance** in factor-model tests to account for the large number of tests that have already been conducted in the profession

- Specifically, **the t-statistic should be at least above 3**, and it will be higher in the future as more tests are conducted

The problem in a nutshell

- If you run one test, the Type I error probability is $\alpha = 5\%$
  - That is the probability of rejecting the null when the null is true
- If you run M independent tests, the probability of at least one **false discovery** – a rejection of the null (a significant factor) when the null is true – is given by the complement to 1 of the probability that none of the M test rejects the null

$$1 - (1 - \alpha)^M \gg \alpha$$

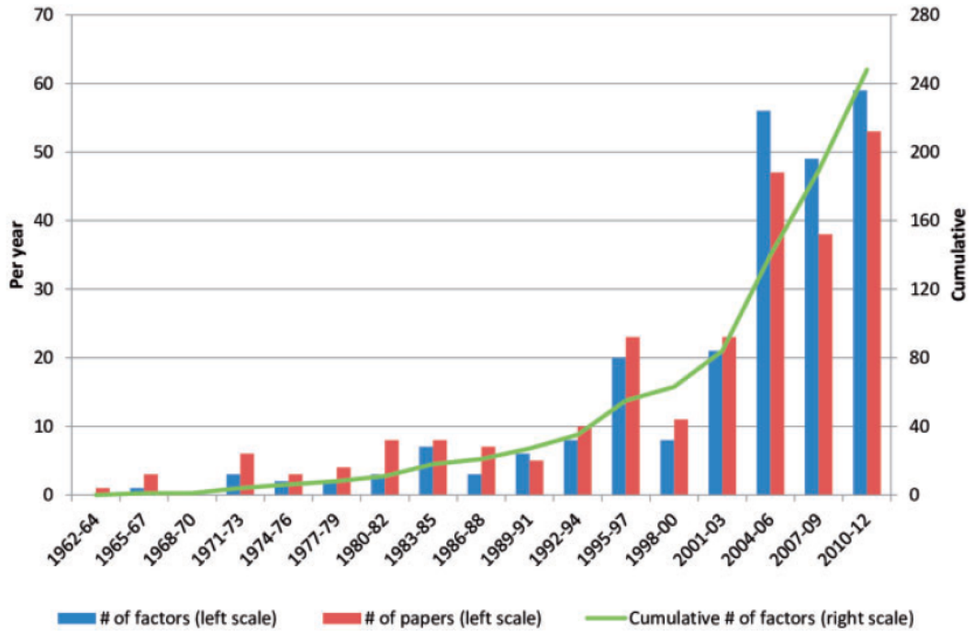
- This probability is increasing in M. E.g. with $M = 100$, $1 - (1 - \alpha)^M = 0.99408$
- Thus, with very high probability the profession will falsely discover factors

Statistical Remedies

- In frequentist statistics ($\neq$ bayesian statistics, see later), the approaches to tackle multiple testing have followed two routes
  1. Control the **family-wise error rate** (FWER): *the probability of at least one false discovery*
     - This method becomes very restrictive as M increases to control the rising number of false negatives as M gets large
     - We have the *Bonferroni* adjustment – set the p-value of each test to α/M – and the *Holm* adjustment – an approach based on the ordering of single tests p-values and the selection of the tests with a p-value below a certain threshold, which depends on the number of tests
  2. Control the **false discovery rate** (FDR): *the expected number of false discoveries over the total discoveries*
     - This approach is more lenient in terms of allowing false discoveries as the number of tests increases
     - We have the *Benjamini, Hochberg, and Yekutieli*'s adjustment, also based on the ordering of p-values

How large is M?

- They consider M equal to the number of factors/anomalies that were published at a given point in time, for a total of 316 (published + WPs) at the end of their sample (2012)
- **Note:** 316 is a lower bound – HLZ estimate that roughly **70% of all tests conducted are never published** (the **file-drawer problem**: only significant results get submitted). Accounting for hidden tests pushes the BHY threshold above 3.0 even under the most lenient correction
- Here is the evolution of published factors



- **What the trend implies:** if the true number of priced factors were finite, the marginal rate of new discoveries should *decline* over time as low-hanging fruit is picked. Instead, the figure shows the opposite – an accelerating pace, from roughly 1 factor/year in the 1980s to 18/year by the late 2000s. This is **consistent with p-hacking**: as data and computing power became cheaper, researchers searched harder and found more, but not because markets became less efficient

Prescriptions for the T-statistics

- They convert the adjusted p-values into adjusted t-statistics
- In 2012, the adjusted t-statistic was between 3 and 3.7
- Assuming the same trend in publication of factors, in 2025, the adjusted t-statistic should be between 3.4 and 4
- Applying this approach back in time, many anomalies (e.g. **SMB, Ivol, E/P**) would be deemed **false discoveries**, while **HML and Momentum would survive**
- In total, **between 80 and 158 (i.e., 27% to 53%) discoveries out of 296 published factors would be deemed false discoveries**

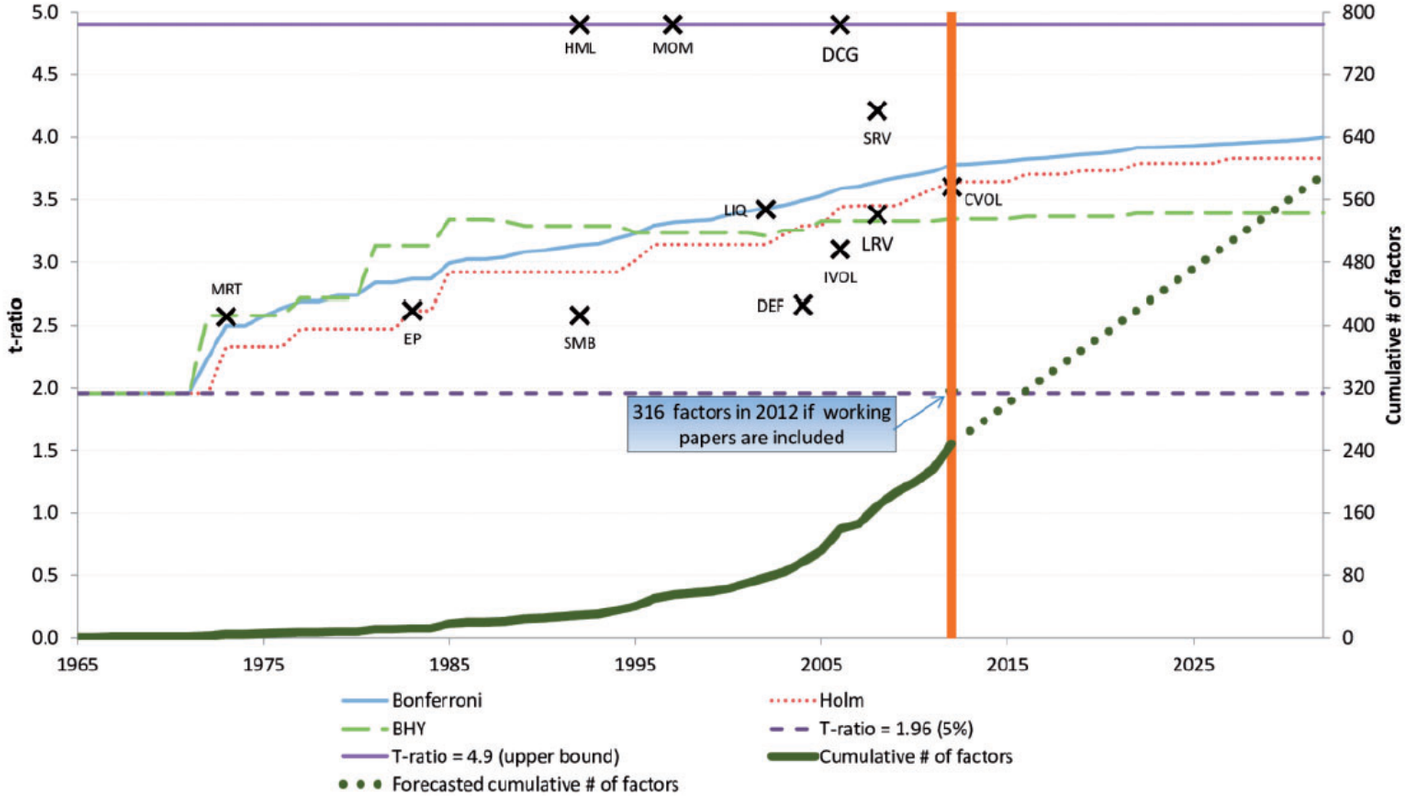


Figure 3
Adjusted *t*-statistics, 1965–2032.
Bonferroni and Holm are multiple testing adjustment procedures that control the family-wise error rate (FWER) and are described in Sections 4.4.1 and 4.4.2, respectively. BHY is a multiple testing adjustment procedure that controls the false discovery rate (FDR) and is described in Section 4.4.3. The green solid curve shows the historical cumulative number of factors discovered, excluding those from working papers. Forecasts (dotted green line) are based on a linear extrapolation. The dark crosses mark selected factors proposed by the literature. They are MRT (market beta; Fama and MacBeth 1973), EP (earnings-price ratio; Basu 1983), SMB and HML (size and book-to-market; Fama and French (1992)), MOM (momentum; Carhart 1997), LIQ (liquidity; Pastor and Stambaugh 2003), DEF (default likelihood; Vassalou and Xing 2004), IVOL (idiosyncratic volatility; Ang et al. 2006); DCG (durable consumption goods; Yogo 2006), SRV and LRV (short-run and long-run volatility; Adrian and Rosenberg 2008), and CVOL (consumption volatility; Boguth and Kuehn 2012). *t*-statistics over 4.9 are truncated at 4.9. For detailed descriptions of these factors, see Table 6 and http://faculty.fuqua.duke.edu/~charvey/Factor-List.xlsx.

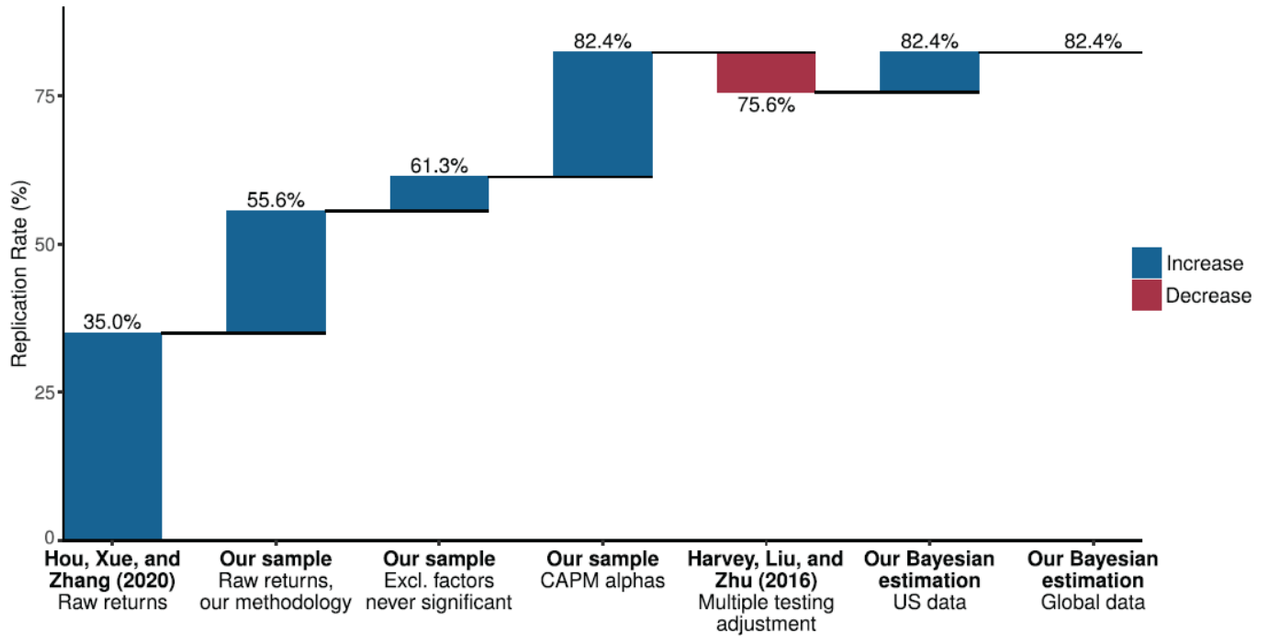
The Limits of P-Hacking (Chen, 2021 JF)

- Chen (2021, JF) is eye-opening on the **limits of p-hacking**
- To produce the 300+ significant factors found in the literature, it would take **hundreds of years** for 10,000 researchers generating 8 fake factors per day
- Intuition: If t-stats are random variables drawn from a N(0,1), the probability of a $|t-stat|>2$ is 0.05, i.e., very low. It takes *a lot* of trying to obtain significance just by chance! Especially considering **t-stats>6 (one in 500 million)**, which are found for dozens of published factors
- The paper does not refute the idea of some spurious factors. **It just resets the prior**: the literature contains genuine signal, so the right question isn't whether factors exist but which ones are robust and non-redundant
- The argument about the prior allows us to segway to the next paper

Jensen, Kelly, and Pedersen (2023, JF)

- Is there a replication crisis in finance?
- Their answer is **NO**
- They show it in three ways:
  1. *Internal Validity*: Replicate the anomalies using a similar methodology to the original studies, while extending the sample to the latest available data
     - Their improvement in replication relative to Hou, Xue, and Zhang (2020) is due to: **longer sample and robust construction choices** (e.g. **monthly holding period**)
  2. *External Validity*: International data (93 countries), post-publication sample
  3. *Bayesian Estimation*: They tackle the Multiple Testing (MT) problem using an **Empirical Bayes Approach** (EB)

- **Anatomy of the replication gap** (their Figure 1): Starting from Hou et al.'s 35%, the rate rises to 56% with a longer sample and consistent construction choices, to 61% after excluding factors never claimed significant in the original papers, and to **82%** once CAPM alphas replace raw returns and the Bayesian correction is applied



Bayesian Statistics in a nutshell

- The idea of Bayesian statistics is that the data provides a reason to update a *prior distribution* of a parameter
- After the updating, one obtains the *posterior distribution* of the parameter
- In our case, a factor's performance is given by

$$f_t = \alpha + \beta r_t^m + \varepsilon_t \tag{6}$$

  with $\varepsilon \sim N\left(0, \sigma^2\right)$
- The prior on alpha is a zero-mean normal distribution

$$\alpha \sim N\left(0, \tau\right)$$

- After using a sample of T observations to obtain an OLS estimate $\hat{\alpha}$, the posterior expected α is

$$E\left(\alpha|\hat{\alpha}\right) = \kappa\hat{\alpha}$$

with updating parameter κ

$$\kappa = \frac{1}{1 + \frac{\sigma^2}{\tau^2 T}} \in (0, 1)$$

Note that:

- The more uninformative the prior is (larger τ), the larger the updating
- The more noisy the data are (larger σ^2), the smaller the updating
- The more data you have (larger T), the larger the updating

- The updating parameter therefore shrinks the posterior towards zero, which is the prior
- The posterior variance is

$$Var\left(\alpha|\hat{\alpha}\right) = \kappa\frac{\sigma^2}{T}$$

- Bottom line: The posterior is shrunk more towards zero (the prior) the more confidence I have in the prior (smaller τ), the smaller the sample I have (T), and the more noisy the data (σ)

Hierarchical Bayesian Model

- Now, let us introduce **multiple testing** (MT) in the form of multiple factors
- Suppose you can obtain **information about alpha from different levels**:
  - Factors in the same cluster, where clusters are: Value, Profitability, etc. They form **13 clusters of factors with average within-group correlation above 50%.** The data are available on their website
  - The same factors across countries
  - Factor-country specific component of the factor
- For illustration, let us say that alpha has a common component c and a factor-specific component w^i

$$\alpha^i = c + w^i$$

with each component having a prior distribution

- You can estimate an alpha for each one of the N factors separately

$$f_t^i = \alpha^i + \beta^i r_t^m + \varepsilon_t^i$$

- Then, you obtain a posterior for each alpha i based on the estimates of the N alphas for the factors (as $N \to \infty$)

$$E\left(\alpha^i | \hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^N\right) = \frac{1}{1 + \frac{\rho\sigma^2}{\tau_c^2 T}} \hat{\alpha}^{\cdot} + \frac{1}{1 + \frac{(1-\rho)\sigma^2}{\tau_c^2 T}} \left(\hat{\alpha}^i - \hat{\alpha}^{\cdot}\right) \tag{7}$$

  with $\hat{\alpha}^{\cdot} = \frac{1}{N} \sum_j \hat{\alpha}^j$ and $\rho = Corr\left(\varepsilon^i, \varepsilon^j\right)$

- All factors are informative about the alpha of factor i because there is a common component of mispricing c. Hence, the first term in Equation (7)

- Then, the individual realization of $\hat{\alpha}^i$ tells us how much we should independently update factor i. Hence, the second term in Equation (7)

- Moreover, you have **two levels of shrinkage** of your estimated $\hat{\alpha}^i$

1. Towards you prior of zero through the updating parameters that are smaller than 1 in general
2. Towards the group mean estimate through the first term in Equation (7)

- Importantly

$$Var\left(\alpha^i|\hat{\alpha}^1,\hat{\alpha}^2,\ldots,\hat{\alpha}^N\right) < Var\left(\alpha^i|\hat{\alpha}^i\right)$$

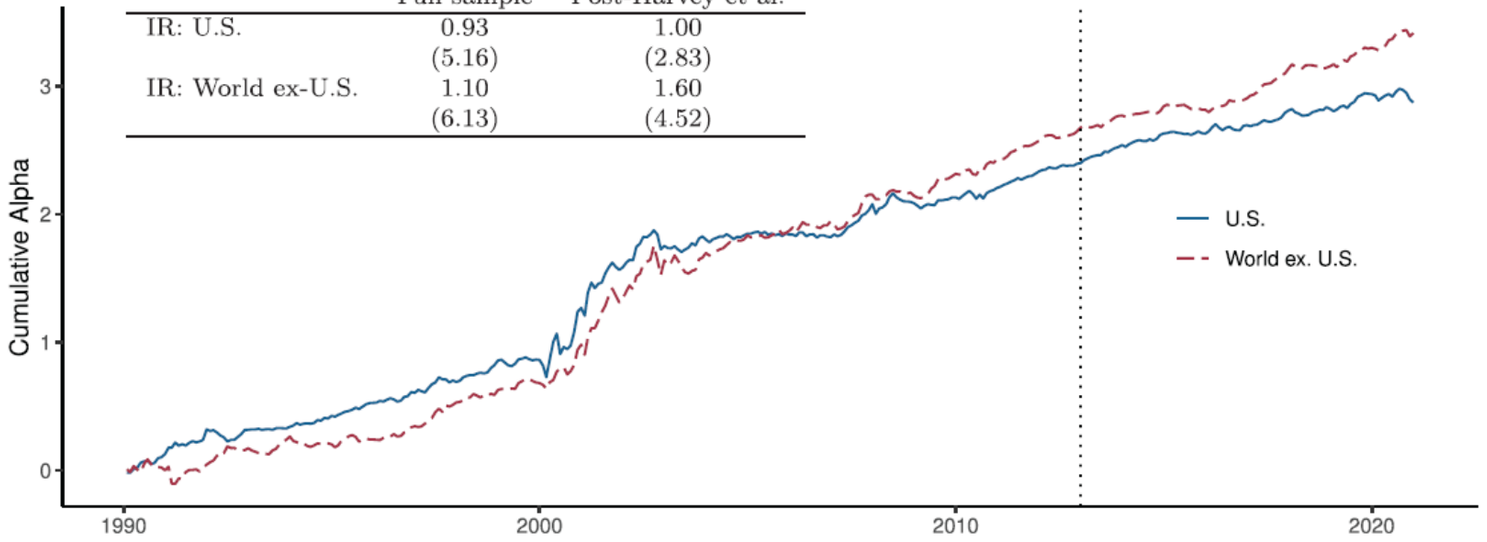
- Intuitively, if you have **many layers that bring information about the same underlying alpha** – e.g., several value factors from different countries or construction methods – which you can use to reduce your posterior uncertainty about the alpha of the factor

- As a result, the Bayesian **confidence intervals are tighter** than single-factor OLS intervals, and also narrower than frequentist multiple-testing-corrected intervals (Harvey, Liu, and Zhu, 2016). Hence, the Bayesian approach **discovers more factors while still controlling the false discovery rate**, leading to higher replication

The Economic Benefits of Less Stringency

- The authors argue that in the end the merits of a statistical procedure to control false discoveries in finance have to be assessed in terms of the **implications for investor performance**
- False discoveries are bad only if they lead to poor performance
- Then, they compare the OOS performance, **measured as Information Ratio (IR)**, of the marginally significant factors, i.e., those that pass the Bayesian threshold but do not pass the more stringent frequentist threshold (Harvey et al., 2016)

	Full sample	Post-Harvey et al.
IR: U.S.	0.93	1.00
	(5.16)	(2.83)
IR: World ex-U.S.	1.10	1.60
	(6.13)	(4.52)



- **The OOS performance is highly satisfactory**, even in the post-Harvey et al (2016) sample
- **Conclusion: we should not talk about a replication crisis** and continue to use these factors

The Virtue of Complexity (Kelly, Malamud, and Zhou, 2024, JF) [AI summary ;-)]

- **Question**: Does adding more and more predictors to a model improve return forecasting, even when the number of predictors vastly exceeds the number of observations?
- **Traditional answer**: No – overfitting destroys out-of-sample performance. Keep models parsimonious
- **KMZ's answer**: Yes – **complexity is a virtue**
  - **Intuition**: if the true return-generating process is complex and unknown, a simple model will always miss signals. **Adding more features captures more of the truth**, and in the overparameterized regime the model automatically self-regularizes by choosing the smallest possible coefficients
  - They call parsimony **"Occam's Blunder"**: simplicity is only optimal if your simple model is correctly specified

– Empirical evidence: using 15 standard macro-financial predictors transformed into up to 12,000 features, trained on only 12 monthly observations, the most complex models achieve the highest out-of-sample Sharpe ratios, and successfully exit the market before 14 of 15 NBER recessions

Seemingly Virtuous Complexity (Nagel, 2025) [AI summary ;-)]

- **Note**: This debate is about **time-series predictability** – whether a model can time the aggregate market. This course is entirely about **cross-sectional predictability** – which stocks outperform others. The two settings are different, and Nagel himself is more conceding on the VoC in the cross-section, where the effective sample size is much larger (thousands of stocks $\times$ hundreds of months)

- **Nagel's critique**: the impressive out-of-sample performance of KMZ's complex models has nothing to do with learning from the data

- **Key observation**: when you train a 12,000-parameter model on only 12 observations, **the model cannot possibly learn anything** – there is simply not enough data. What it does instead is mechanically construct a well-known simple strategy

- **What the complex model actually is**: a **volatility-timed momentum strategy** – it bets on the market going up when recent returns have been positive and volatility has been low. No machine learning is actually taking place

- **Why persistence generates momentum – in formulas**

  Step 1: minimum-norm ridge regression produces a weighted average of past returns. With $P > T$, the minimum-norm solution gives forecast:

$$\hat{r}_{t+1} = \sum_{s=1}^{T} w_s \cdot r_{s+1}$$

  where the weight on past observation s depends on the **similarity** between today's predictor vector x_t and the past predictor vector x_s:

$$w_s \propto k(x_t, x_s) = \exp\left(-\frac{\gamma^2}{2}\|x_t - x_s\|^2\right)$$

  Past returns that followed a predictor environment similar to today get high weight; dissimilar episodes get low weight. **This is how the predictive content of x is supposed to enter**: Suppose this month is a high-dividend yield period. The predicted return will be similar to returns in previous episodes with high dividend yield.

  Step 2: persistence kills the signal. With persistent predictors, write $x_s \approx x_t - (t-s)\delta$ where δ is the monthly drift. Then $\|x_t - x_s\|^2 \approx (t-s)^2\|\delta\|^2$ and:

$$w_s \propto \exp\left(-\lambda(t-s)^2\right), \qquad \lambda = \frac{\gamma^2\|\delta\|^2}{2}$$

The weight now depends **on the lag** $t - s$, and on the extent of variability in the serie δ. With T=12 monthly observations, all past predictor vectors look nearly identical to today's – the dividend yield barely moved. The kernel cannot tell a high-yield episode from a low-yield one. Weights collapse to pure recency weights, and the forecast becomes a **recency-weighted average of past returns: time-series momentum**. The predictive content of x is never identified because you don't have enough variation of regimes in your data

- **Evidence**:
  - After controlling for the simple momentum-volatility strategy, the 12,000-parameter model adds **zero** additional alpha
  - When Nagel artificially reverses the autocorrelation in returns (making momentum a bad strategy), the complex model still builds a momentum strategy and now **loses money**. A model that truly learned from the data would have adapted
  - Randomly scrambling the predictors barely changes the model's performance – the actual information in the predictors is nearly irrelevant
- **Is Nagel attacking the theory? No** – he explicitly accepts KMZ's mathematical propositions. His challenge is subtler: **the theory assumes that additional features carry some genuine signal.**

His experiments show that in the $T = 12$ setup, the features carry **effectively zero signal** – performance is the same whether you use the real predictors or randomly scrambled ones. So the key assumption of the theorem is violated in the very setup used to illustrate it. The theory is correct; it just does not apply to this case

- **Bottom line**: Complexity may be virtuous in principle and with enough data; with 12 observations it is an illusion
  - This does not mean ML has no role in finance. Cross-sectional prediction — forecasting which stocks outperform others — offers thousands of stocks times hundreds of months, a much richer environment where genuine learning is plausible
  - The lesson: import ML tools selectively, always asking whether the effective sample size is large enough to support the complexity being claimed

What is not included in these notes

- For reasons of space, I could not discuss very important work, which you should look at if you are interested in working in this area
  - Studies of how expectations are formed and impact asset prices. E.g., Bordalo, Gennaioli, La Porta, and Shleifer (2019), and other papers by the same authors. Adam and Nagel (2023): for a survey
  - Machine learning and Asset Pricing. Start from Nagel's (2021) textbook and Kelly and Xiu (2023)'s Financial Machine Learning
  - Asset pricing from the point of view of prices, not expected returns. Cho and Polk (2024)
  - Structural models providing explicit formulation of the SDF and testing them. E.g. Kogan and Papanikolaou (2013)