Section 2: The factor zoo

Ralph S.J. Koijen   Stijn Van Nieuwerburgh*

September 17, 2021

*Koijen: University of Chicago, Booth School of Business, NBER, and CEPR. Van Nieuwerburgh: Columbia Business School, CEPR, and NBER. If you find typos, or have any comments or suggestions, then please let us know via ralph.koijen@chicagobooth.edu or svnieuwe@gsb.columbia.edu.

1. Basic structure of the notes

- High-level summary of theoretical frameworks to interpret empirical facts.
- Per asset class, we will discuss:
  1. Key empirical facts in terms of prices (unconditional and conditional risk premia) and asset ownership.
  2. Interpret the facts using the theoretical frameworks.
  3. Facts and theories linking financial markets and the real economy.
  4. Active areas of research and some potentially interesting directions for future research.
- The notes cover the following asset classes:
  1. Equities (weeks 1-5).
     - Predictability and the term structure of risk (week 1)
     - The Cross-section and the factor zoo (week 2)
     - Intermediary-based asset pricing (week 3)
     - Production-based asset pricing (week 4)
     - Asset pricing via demand systems (week 5)
  2. Mutual funds and hedge funds (week 6).
  3. Options and volatility (week 7).
  4. Government bonds (week 8).
  5. Corporate bonds and CDS (week 9).
  6. Currencies and international finance (week 10).
  7. Commodities (week 11).
  8. Real estate (week 12).

2. The Cross-Section of Stock Returns

2.1. Cross-sectional Predictability

- Instead of predicting the return on the aggregate stock market, there is a large literature that studies cross-sectional stock return predictability.

- The typical procedure is:

  1. Sort stocks on a characteristic into quintile or decile portfolios and document a pattern in average returns.
  2. Construct a long-short strategy that buys the top quintile or decile and shorts the bottom decile.
  3. The factor constructed in step 2. is then used as a factor to explain average returns.

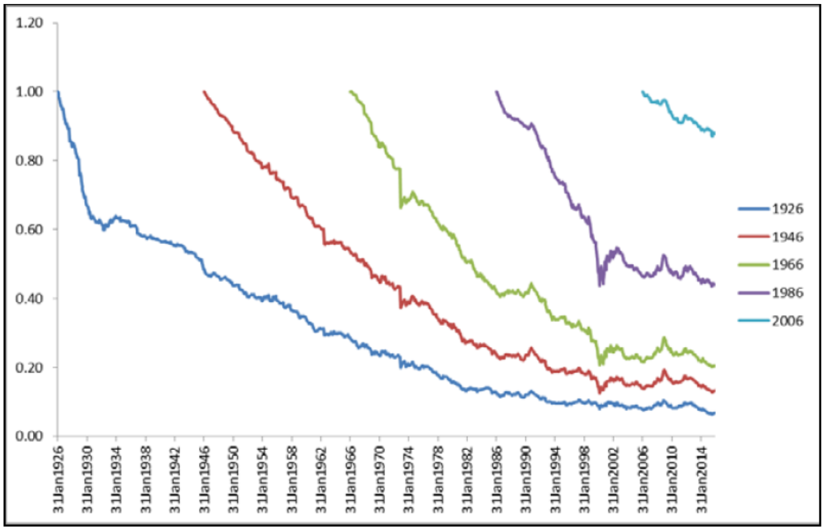
- By forming a long-short strategy, we "net out" some of the passive exposures, for instance, to the market that would arise from a long-only portfolio.

- This argument works as long as, for instance, market betas and the characteristic you sort on are not highly correlated.

- As an alternative to portfolio sorts, you can show that a characteristic predicts returns in the cross-section using the Fama-MacBeth procedure. More on this later.

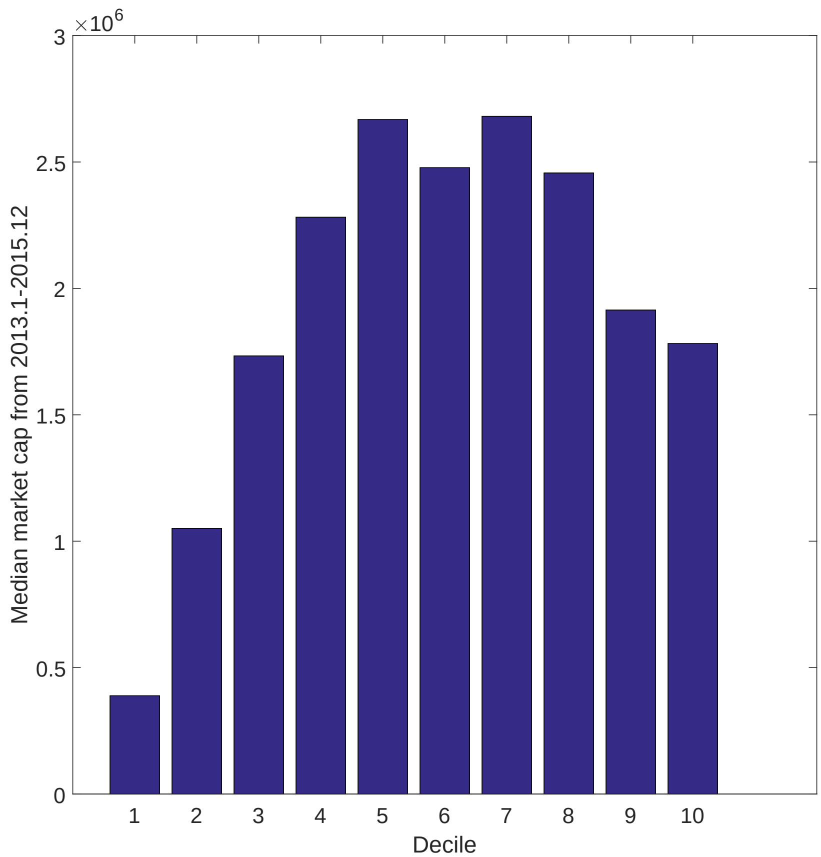
- Typical empirical implementation of an investment strategy for U.S. equities:
  - Equity prices come from CRSP and accounting data is from Compustat. CRSP-Compustat merged provides a variable PERMNO that you can use to merge the data.
  - Accounting data is released to the public with a lag. It is common practice to lag the accounting data by 6 months (sometimes 3 months) to ensure that the accounting data are indeed available to investors at the time of portfolio formation.
  - Most research uses stocks listed on the AMEX, NASDAQ, and NYSE. Sometimes papers impose a minimum price of $1 or $5 to avoid looking at penny stocks.
  - As most firms have their fiscal year-end in December, it is common practice to sort portfolios in June and then track the performance of the portfolios for the next 12 months.
  - To sort stocks into portfolios, we typically use the characteristics of the NYSE stocks, which tend to be larger firms, to determine which stock goes into which portfolio.
  - *Within* each of the portfolios, you can either value-weight or equally-weight the stocks. The results are typically stronger for equally-weighted returns as anomalies tend to be more pronounced for smaller stocks. However, value-weighting arguably leads to economically more meaningful results.
  - If a firm defaults, it is important to use the delisting return when available. Although imperfect, this is better than forgetting about delisting returns altogether. See Shumway (1997) for a discussion of delisting returns.

- Note 1: Entry and exit in the stock market matters, see for instance Pedersen (2018), although this issue has not received as much attention:



[3] This figure reports the total market value the buy-and-hold strategy as a fraction of the total market value of all shares, adjusting for stock splits by assuming that the passive investor is treated like other investors in any stock split. While some readers might be surprised that doing nothing is not enough to be passive, others may be surprised that the investor from 1926 continues to hold as much as 10% of the market today – this is due to old giants like Standard oil, GE, Chevron, and Coca Cola.

- Note 2: Even when quintile or decile portfolios are value-weighted, this does not mean that there is the same fraction of market cap in each of the quintiles or deciles. For instance, the median market cap in various momentum deciles (sort on the price change over the last 12-2 months):



- Even if stocks in the extreme deciles are mispriced, how large is the overall mispricing at an aggregate level?

- Looking at quintile and decile portfolios, even when value-weighted within the portfolio, may be misleading.

- See Hou, Xue, and Zhang (2020) for more on the role of firm size and anomaly returns.

2.2. *Basic Equity Return Factors*

- Main cross-sectional predictors that have been studied in the literature

  - Market beta.
  - Market capitalization (“size”).
  - Book-to-market (“value”).
  - Lagged price changes (“Momentum”).
  - Investment / asset growth.
  - Profitability.
  - Liquidity.

2.2.1. Market beta

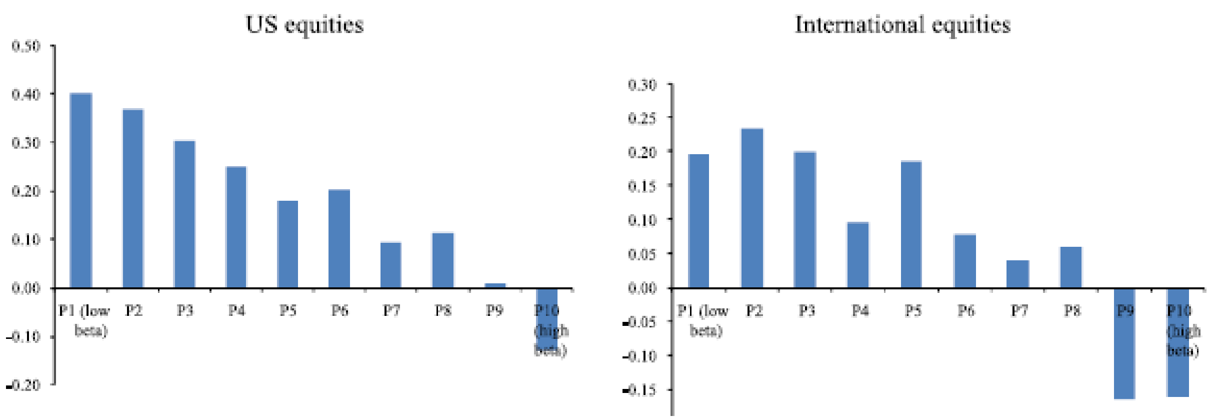
- The CAPM obviously motivates market beta as a variable to sort stocks on. The Security Market Line:

$$E_t[R^{i,e}_{t+1}] = \beta^i E_t[R^{m,e}_{t+1}]$$

Empirically implemented as the Security Characteristic Line:

$$R^{i,e}_{t+1} = \alpha^i + \beta^i R^{m,e}_{t+1} + e^i_{t+1}$$

- However, the security market line appears to be too flat.
- For a recent summary of the facts, see Frazzini and Pedersen (2014). These are the CAPM alphas of beta-sorted portfolios:



- If we simply form a long-short portfolio of low-beta minus high-beta stocks, the portfolio is (by design) not beta-neutral.

- Frazzini and Pedersen (2014) construct a beta-neutral portfolio

$$R_{t+1}^{BAB} = \frac{1}{\beta_L} R_{L,t+1}^e - \frac{1}{\beta_H} R_{H,t+1}^e, \tag{1}$$

  where $R_{L,t+1}^e$ is the excess return on the low-beta portfolio and $R_{H,t+1}^e$ the excess return on the high-beta portfolio.

- This Betting-Against-Beta factor, which is close to market-neutral, earns high Sharpe ratios in most countries and asset classes.

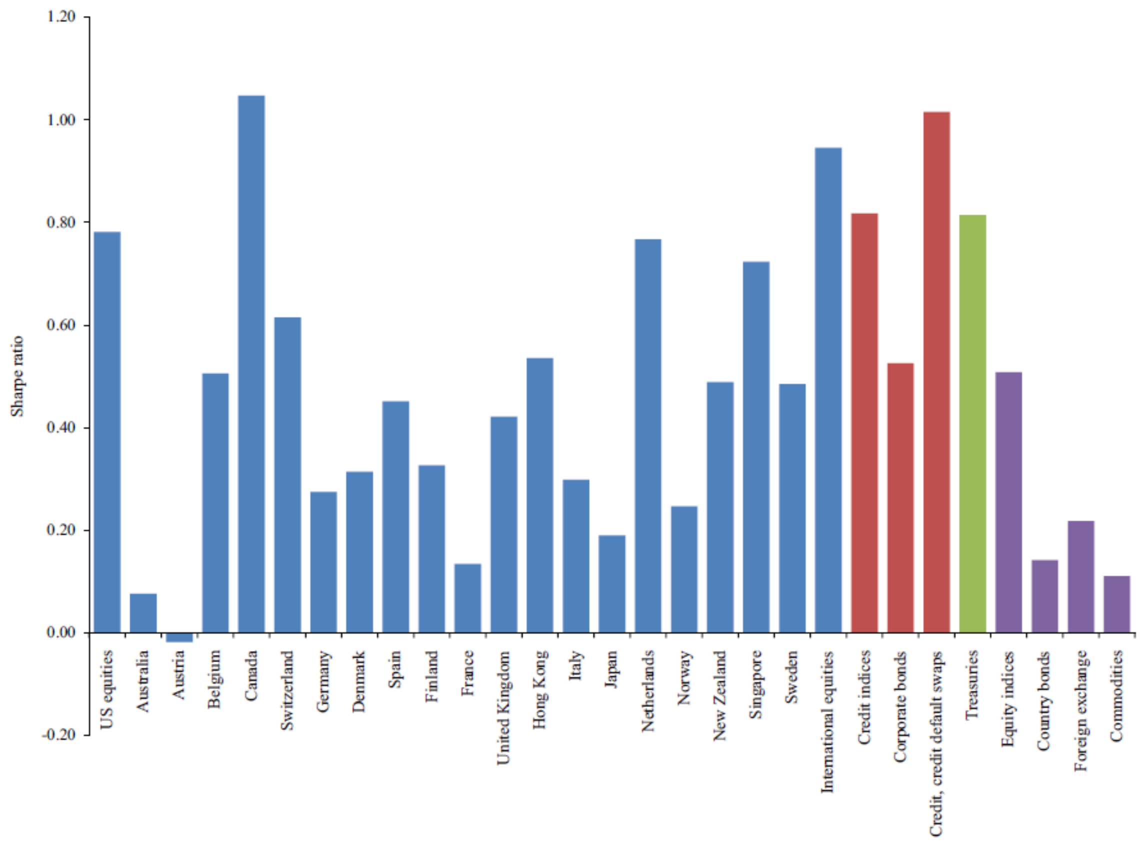


Fig. 2. Betting against beta (BAB) Sharpe ratios by asset class. This figures shows annualized Sharpe ratios of BAB factors across asset classes. To construct the BAB factor, all securities are assigned to one of two portfolios: low beta and high beta. Securities are weighted by the ranked betas and the portfolios are rebalanced every calendar month. Both portfolios are rescaled to have a beta of one at portfolio formation. The BAB factor is a self-financing portfolio that is long the low-beta portfolio and shorts the high-beta portfolio. Sharpe ratios are annualized.

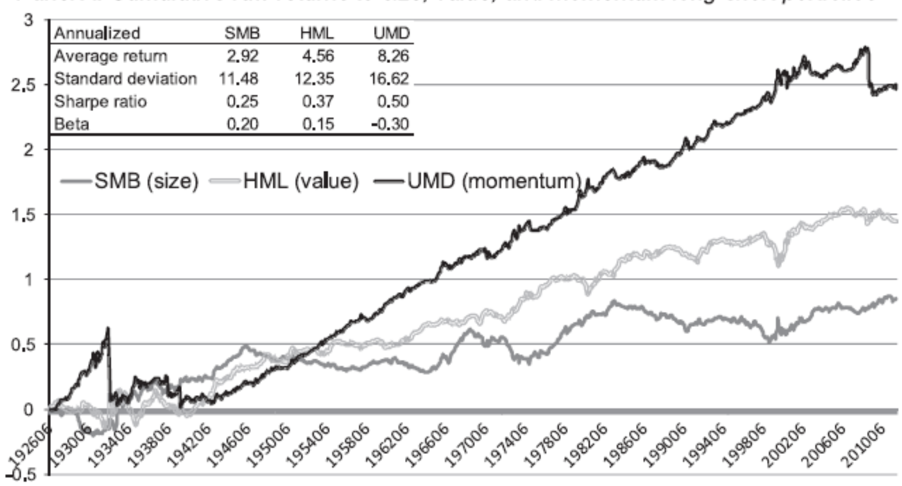
2.2.2. Size

- The size anomaly goes back to Banz (1981), and this is the first major challenge to the CAPM. Small stocks earn positive CAPM alphas.

- Size factor SMB: Small-Minus-Big portfolio

- The cumulative return on size, value, and momentum

Panel A: Cumulative raw returns to size, value, and momentum long-short portfolios



- The size factor has not done particularly well since its discovery in early 1980s.

- Summary from Israel and Moskowitz (2013):

Table 2
Decile portfolios based on size, value, and momentum from July 1926 to December 2011.

Reported are the average raw returns in excess of the one-month T-bill rate, Sharpe ratios, and CAPM alphas of value-weighted decile portfolios formed on size, value (BE/ME), and momentum (past 12-month return, skipping the most recent month) over the period July 1926 (January 1927 for momentum) to December 2011. The difference between Deciles 10 and 1 (10-1) is also reported along with the differences between the average of Deciles 9 and 10 and the average of Deciles 1 and 2 (9-2), the average of Deciles 8 through 10 and the average of Deciles 1 through 3 (8-3), and the average of Deciles 7 through 10 and the average of Deciles 1 through 4 (7-4).

	Decile portfolios (value-weighted)										Differences			
	1	2	3	4	5	6	7	8	9	10	10-1	9-2	8-3	7-4
Size														
Raw excess	13.66	11.56	11.54	10.96	10.49	10.42	9.81	9.11	8.43	6.80	−6.86	−4.99	−4.14	−3.39
Sharpe	0.39	0.37	0.41	0.42	0.42	0.43	0.43	0.42	0.41	0.38	−0.26	−0.24	−0.24	−0.23
Alpha	2.97	1.26	1.63	1.67	1.30	1.47	1.25	0.90	0.54	−0.06	−3.03	−1.87	−1.49	−1.22
t-statistic	(1.22)	(0.71)	(1.18)	(1.38)	(1.34)	(1.83)	(1.85)	(1.67)	(1.29)	(−0.19)	(−1.15)	(−0.89)	(−0.86)	(−0.83)

- The size effect has also been linked to the January effect, tax-loss selling, and window dressing.

- The SMB factor has had an annualized return of -0.2% between January 2010 and June 2020. Compared with a 12.5% excess return on the overall stock market, and compared with an average SMB return of 2.7% from 1926.07-2009.12.

2.2.3. Value

- In case of value, we sort on the book-to-market ratio.
- HML factor = Long value stocks (high BM), short growth stocks (low BM).
- The value premium is larger for small firms. Again from Israel and Moskowitz (2013):

	Smallest				Largest	
	Size 1	Size 2	Size 3	Size 4	Size 5	Size 1-Size 5
VALUE						
Returns						
5-1 spread	11.22	7.28	5.42	4.24	3.70	7.13
	(3.87)	(3.88)	(2.96)	(1.93)	(1.90)	(2.10)
Long side	16.45	14.26	13.48	12.36	11.01	6.07
	(4.58)	(4.34)	(4.17)	(3.67)	(3.86)	(2.74)
Percent long side	146.5	195.9	248.8	291.3	297.6	
Long=short (*t*-statistic)	(2.81)	(3.50)	(3.66)	(3.81)	(4.01)	
Alphas						
5-1 spread	12.99	6.38	4.63	1.54	2.19	10.58
	(4.52)	(3.41)	(2.53)	(0.74)	(1.14)	(3.21)
Long side	6.15	4.15	3.26	1.73	1.97	4.31
	(2.78)	(2.38)	(2.05)	(1.04)	(1.21)	(1.97)
Percent long side	47.4	65.1	70.4	112.9	89.9	
Long=short (*t*-statistic)	(0.15)	(0.67)	(0.89)	(1.18)	(1.15)	

- The HML factor has done very poorly in the last decade. Annualized returns between January 2010 and June 2020 for HML were -5.6% compared to +5.1% for 1926.07-2009.12.
- Normally, small value stocks outperform but large growth companies did much better in the last decade driven by FAANG (Facebook, Amazon, Apple, Netflix, Google; also Microsoft).

- Value is related to long-term reversals, documented by De Bond and Thaler (1985).
- Long-term reversal sorts stocks on their returns between months $t-36$ and $t-12$. Long the losers, short the winners.

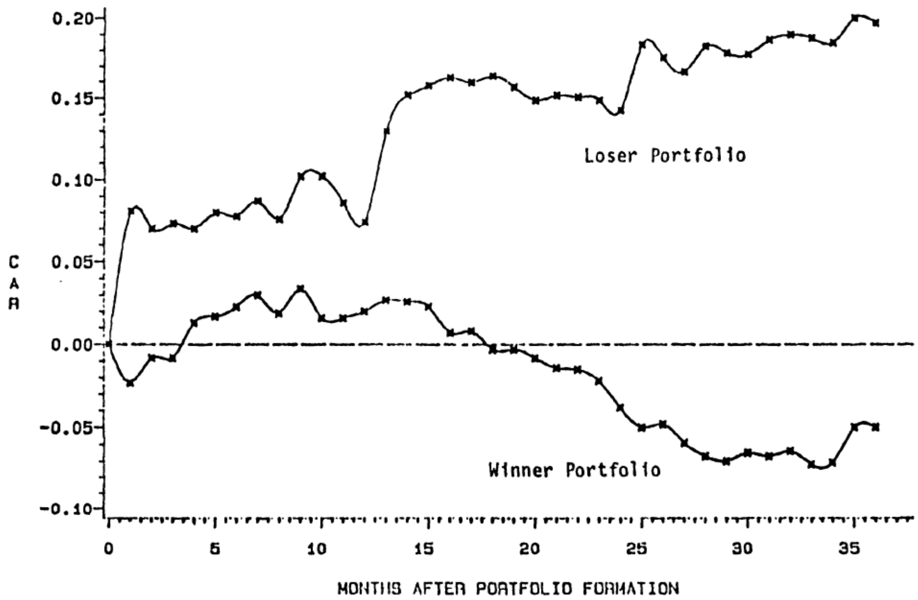


Figure 1. Cumulative Average Residuals for Winner and Loser Portfolios of 35 Stocks (1–36 months into the test period)

2.2.4. Momentum

- In case of momentum, we sort on the price change between months $t-12$ and $t-2$. Long the winners, short the losers.
- The most recent month is excluded as there are short-term reversals (between $t-1$ and t).
- Again, from Israel and Moskowitz (2013), where the columns correspond to size (final column is a long-short strategy across column 5 and column 1):

MOMENTUM						
Returns						
5-1 spread	10.87	12.99	11.53	10.79	7.49	3.42
	(4.50)	(6.22)	(4.76)	(4.08)	(2.95)	(1.56)
Long side	18.76	17.17	15.61	14.98	10.98	7.79
	(5.59)	(5.71)	(5.89)	(6.09)	(4.95)	(3.35)
Percent long side	172.6	132.1	135.3	138.9	146.6	
Long=short (t-statistic)	(3.72)	(3.29)	(3.36)	(3.45)	(2.81)	
Alphas						
5-1 spread	13.12	15.30	14.48	14.19	10.24	2.88
	(5.59)	(7.66)	(6.32)	(5.72)	(4.23)	(1.31)
Long side	9.30	7.89	7.26	7.17	3.92	5.37
	(4.47)	(5.13)	(5.71)	(6.24)	(3.83)	(2.40)
Percent long side	70.9	51.6	50.1	50.5	38.3	
Long=short (t-statistic)	(1.34)	(0.17)	(0.02)	(0.09)	(1.78)	

- Observations:
  - Momentum returns and CAPM alphas are large one average.
  - Momentum returns are not as strongly related to size.
  - In forming a momentum portfolio, only information on prices is required, which makes the momentum anomaly all the more challenging to explain.
- Momentum (UMD factor) had annualized returns of 4.1% from January 2010–June 2020, only half as large as the 8.4% return from 1927.01-2009.12.

2.2.5. Investment

- Investment is typically defined as capital expenditures.
- However, in the recent literature, changes in assets are labeled investment.
- Asset growth strongly predicts returns in the cross-section, where firms with *high* asset growth have *low* average returns.
- The main fact has been documented in Cooper, Gulen, and Schill (2008):

Panel B.2: Value-Weighted Portfolio Average Monthly Raw Returns

	Asset Growth Deciles											
YEAR	1(Low)	2	3	4	5	6	7	8	9	10(High)	Spread (10-1)	t(spread)
−5	0.0121	0.0123	0.0117	0.0129	0.0142	0.0146	0.0165	0.0207	0.0243	0.0271	0.0150	7.55
−4	0.0114	0.0109	0.0119	0.0131	0.0128	0.0146	0.0172	0.0202	0.0288	0.0307	0.0193	9.15
−3	0.0064	0.0085	0.0100	0.0123	0.0143	0.0151	0.0157	0.0212	0.0279	0.0357	0.0292	10.92
−2	0.0062	0.0083	0.0090	0.0116	0.0135	0.0149	0.017	0.0206	0.0266	0.0396	0.0334	12.86
−1	0.0223	0.0175	0.0153	0.0146	0.0147	0.0141	0.0153	0.0177	0.0192	0.0230	0.0007	0.28
1	0.0148	0.0124	0.0122	0.0116	0.0100	0.0100	0.0102	0.0092	0.0077	0.0043	−0.0105	−5.04
2	0.0133	0.0126	0.0125	0.0101	0.0109	0.0102	0.0098	0.0097	0.0097	0.0065	−0.0068	−3.39
3	0.0169	0.0137	0.0141	0.0126	0.0102	0.0112	0.0116	0.0105	0.0116	0.0116	−0.0053	−2.82
4	0.0132	0.0107	0.012	0.0109	0.0114	0.0103	0.0103	0.0123	0.0111	0.0120	−0.0012	−0.61
5	0.0128	0.0133	0.0121	0.0123	0.0103	0.01	0.0107	0.0113	0.013	0.0126	−0.0002	−0.11
	Cumulative Return											
[−5,−1]	1.0449	0.9918	1.0078	1.2375	1.3631	1.4788	1.7985	2.5321	3.9221	6.4272	5.3822	4.78
[1, 5]	1.2879	1.1133	1.1305	1.0038	0.931	0.8934	0.9352	0.9056	0.9458	0.7911	−0.4967	−4.25

- Fama and French call their investment factor CMA. It goes long conservative (i.e., low investment) stocks and short aggressive (high investment) stocks.

2.2.6. Profitability

- Profitability is defined slightly differently in the main papers:
  1. Novy-Marx (2013).
  2. Hou, Xue, and Zhang (2015).
  3. Fama and French (2015).
- The facts below are based on Novy-Marx (2013) using gross profits/assets

Table 2
Excess returns to portfolios sorted on profitability.
This table shows monthly value-weighted average excess returns to portfolios sorted on gross profits-to-assets [(REVT − COGS)/AT], employing NYSE breakpoints, and results of time series regressions of these portfolios' returns on the Fama and French factors [the market factor (MKT), the size factor small-minus-large (SMB), and the value factor high-minus-low (HML)], with test-statistics (in square brackets). It also shows time series average portfolio characteristics [portfolio gross profits-to-assets (GP/A), book-to-market (B/M), average firm size (ME, in millions of dollars), and number of firms (n)]. Panel B provides similar results for portfolios sorted on book-to-market. The sample excludes financial firms (those with one-digit standard industrial classification codes of six) and covers July 1963 to December 2010.

		Alphas and three-factor loadings				Portfolio characteristics			
Portfolio	r^e	α	MKT	SMB	HML	GP/A	B/M	ME	n
Panel A: Portfolios sorted on gross profits-to-assets									
Low	0.31	−0.18	0.94	0.04	0.15	0.10	1.10	748	771
	[1.65]	[−2.54]	[57.7]	[1.57]	[5.87]				
2	0.41	−0.11	1.03	−0.07	0.20	0.20	0.98	1,100	598
	[2.08]	[−1.65]	[67.5]	[−3.13]	[8.51]				
3	0.52	0.02	1.02	−0.00	0.12	0.30	1.00	1,114	670
	[2.60]	[0.27]	[69.9]	[−0.21]	[5.42]				
4	0.41	0.05	1.01	0.04	−0.24	0.42	0.53	1,114	779
	[1.94]	[0.83]	[70.6]	[1.90]	[−11.2]				
High	0.62	0.34	0.92	−0.04	−0.29	0.68	0.33	1,096	938
	[3.12]	[5.01]	[58.3]	[−2.03]	[−12.3]				
High–low	0.31	0.52	−0.03	−0.08	−0.44				
	[2.49]	[4.49]	[−0.99]	[−2.15]	[−10.8]				
Panel B: Portfolios sorted on book-to-market									
Low	0.39	0.13	0.98	−0.09	−0.39	0.43	0.25	1,914	965
	[1.88]	[2.90]	[90.1]	[−5.62]	[−23.9]				
2	0.45	−0.02	0.99	0.05	0.04	0.31	0.54	1,145	696
	[2.33]	[−0.29]	[78.1]	[2.61]	[2.23]				
3	0.56	0.03	0.96	0.04	0.22	0.26	0.79	849	640
	[2.99]	[0.53]	[63.5]	[2.09]	[9.71]				
4	0.67	−0.00	0.96	0.10	0.53	0.21	1.12	641	655
	[3.58]	[−0.03]	[74.8]	[5.66]	[27.1]				
High	0.80	0.07	1.01	0.25	0.51	0.21	5.47	367	703
	[3.88]	[1.04]	[60.7]	[10.7]	[20.5]				
High–low	0.41	−0.06	0.03	0.34	0.91				
	[2.95]	[−0.71]	[1.44]	[12.0]	[30.0]				

2.2.7. Liquidity

- Huge literature on liquidity and liquidity risk in finance.
- Pastor and Stambaugh (2003) study whether stocks that are exposed to *aggregate* liquidity shocks earn higher expected returns.
- Measurement challenge: How to measure liquidity risk?
- In theory, liquidity is the slope of the demand curve. In less liquid markets, the residual demand curve is steeper.
- Hence, trades move prices temporarily from "fundamental" value, and then gradually return. The paper focuses on temporary price changes associated with order flow.
- For a classic paper on liquidity, see Campbell, Grossman, and Wang (1993).

- The three main steps to construct the Pastor and Stambaugh aggregate liquidity factor:

  1. Regress returns on lagged returns and dollar volume times the sign of lagged returns within a month for a given stock

     $$R^e_{i,d+1,t} = \theta_{i,t} + \phi_{i,t} R^e_{i,d,t} + \gamma_{i,t} sign(R^e_{i,d,t}) v_{i,d,t} + \epsilon_{i,d+1,t}, d = 1, \ldots, D,$$

     where $R^e_{i,d,t} = R_{i,d,t} - R_{m,d,t}$ with $R_{m,d,t}$ the value-weighted market return.

     $\gamma_{i,t}$ measures how price changes relate to trading the previous period. If there was a significant amount of trade and the return was negative, we expect prices to partially recover over the next day. Hence, we expect $\gamma_{i,t} < 0$.

  2. We are interested in liquidity shocks, and hence would like to compute innovations. A problem is that dollar volume trends over time. First differences are constructed as:

     $$\Delta \widehat{\gamma}_t = \frac{m_t}{m_1} N^{-1} \sum_{i=1}^{N} (\widehat{\gamma}_{i,t} - \widehat{\gamma}_{i,t-1}),$$

     where where m_t is the total dollar value of the stock market at the end of month $t-1$ of the stocks included in the average in month t, and month 1 corresponds to August 1962, the beginning of the sample.

  3. Aggregate liquidity shocks are computed as

     $$\Delta \widehat{\gamma}_t = a + b \Delta \widehat{\gamma}_{t-1} + c \frac{m_{t-1}}{m_1} \widehat{\gamma}_{t-1} + u_t,$$

     where u_t is the liquidity shock that is used in the empirical analysis. High u_t is a high **illiquidity** shock.

- Form portfolios of stocks sorted by exposure to u_t
- Value-weighted portfolio alphas by sorting on illiquidity betas:

TABLE 4
ALPHAS OF VALUE-WEIGHTED PORTFOLIOS SORTED ON PREDICTED LIQUIDITY BETAS

	Decile Portfolio										
	1	2	3	4	5	6	7	8	9	10	10–1
	A. January 1966–December 1999										
CAPM alpha	−5.16	−1.88	−.66	−.07	−1.48	1.48	1.22	1.38	1.68	1.24	6.40
	(−2.57)	(−1.24)	(−.56)	(−.08)	(−1.80)	(1.93)	(1.52)	(1.72)	(1.93)	(1.01)	(2.54)
Fama-French alpha	−6.05	−3.36	−2.15	−1.23	−2.10	.78	.86	1.41	1.90	3.18	9.23
	(−3.77)	(−2.47)	(−1.93)	(−1.37)	(−2.61)	(1.08)	(1.11)	(1.76)	(2.22)	(2.82)	(4.29)
Four-factor alpha	−5.11	−1.66	−1.02	−.76	−1.61	.91	.76	1.55	1.34	2.36	7.48
	(−3.12)	(−1.23)	(−.91)	(−.83)	(−1.96)	(1.22)	(.96)	(1.88)	(1.54)	(2.06)	(3.42)
	B. January 1966–December 1982										
CAPM alpha	−2.26	1.63	.54	.67	−3.09	1.44	.61	1.78	1.43	−.93	1.34
	(−.81)	(.76)	(.31)	(.50)	(−2.69)	(1.29)	(.54)	(1.46)	(1.14)	(−.52)	(.36)
Fama-French alpha	−7.32	−2.22	−1.80	−.75	−3.29	1.03	.20	1.91	2.32	1.18	8.50
	(−3.36)	(−1.23)	(−1.13)	(−.59)	(−2.85)	(.95)	(.17)	(1.56)	(1.86)	(.71)	(2.77)
Four-factor alpha	−6.43	−.25	−.22	−.03	−2.46	1.09	.31	2.89	1.67	−.22	6.21
	(−2.82)	(−.13)	(−.13)	(−.02)	(−2.05)	(.95)	(.25)	(2.28)	(1.28)	(−.13)	(1.95)
	C. January 1983–December 1999										
CAPM alpha	−8.01	−5.33	−1.76	−1.01	.20	1.55	1.74	.70	1.81	3.38	11.39
	(−2.76)	(−2.49)	(−1.08)	(−.77)	(.17)	(1.46)	(1.54)	(.67)	(1.47)	(1.98)	(3.36)
Fama-French alpha	−5.23	−5.08	−2.69	−1.80	−.82	.37	.89	.76	1.25	5.51	10.74
	(−2.23)	(−2.46)	(−1.67)	(−1.41)	(−.72)	(.38)	(.89)	(.72)	(1.05)	(3.51)	(3.53)
Four-factor alpha	−4.43	−3.72	−1.94	−1.52	−.63	.53	.70	.47	.84	5.06	9.49
	(−1.88)	(−1.85)	(−1.21)	(−1.17)	(−.54)	(.54)	(.69)	(.44)	(.70)	(3.20)	(3.12)

NOTE.—See the note to table 3. The table reports the decile portfolios' postranking alphas, in percentages per year. The alphas are estimated as intercepts from the regressions of excess portfolio postranking returns on excess market returns (CAPM alpha), on the Fama-French factor returns (Fama-French alpha), and on the Fama-French and momentum factor returns (four-factor alphas). The t-statistics are in parentheses.

- The alphas are particularly large relative to the Fama and French model and the 4-factor model.
- The alphas appear to be slightly larger in the second part of the sample.

- Interpreting reversals as measures of liquidity, means that the profits of reversal strategies tell you when the benefits of liquidity provision are particularly high. This idea is explored in Nagel (2012).

- The benefits of liquidity provision are strongly correlated with the VIX.

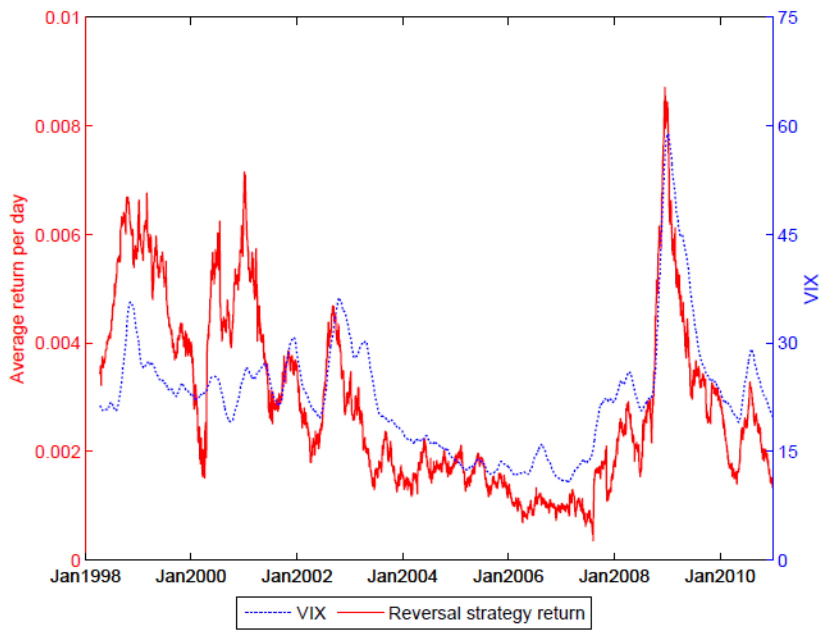


Figure 1: 3-month moving averages of daily return-reversal strategy returns and the CBOE S&P500 implied volatility index (VIX). Each day t, the reversal strategy returns are calculated as the average of returns from five reversal strategies that weight stocks proportional to the negative of market-adjusted returns on days $t-1$, $t-2$, ..., $t-5$, with weights scaled to add up to \$1 short and \$1 long. Returns are calculated from daily CRSP closing transaction prices, and returns are hedged against conditional market factor exposure.

- Classic papers on liquidity in equity markets
  - Measuring price impact: Hasbrouck (1991).
  - Alternative measure of liquidity: Amihud (2002).
  - CAPM with liquidity effects: Acharya and Pedersen (2005).

2.3. *Time-variation in Cross-sectional Predictability*

- In addition to thinking about predictability in the time series and the cross-section separately, there is also a literature studying the predictability of anomaly portfolios.
- Cohen, Polk, and Vuolteenaho (2003) propose a predictor of the value premium, the "value spread," which depends on the spread in valuation (B/M) ratios of value and growth stocks.

Table V

HML return predictability

This table reports regressions predicting the simple return on the HML portfolio. Each regression contains a constant and the HML value spread. In addition, some of the regressions contain combinations of the following variables: the lagged ROE of the HML portfolio, market BE/ME, default yield spread (Moody's BAA less AAA corporate-bond yield), and an interaction term between the value spread and the median BE/ME of the market. The HML value spread is defined as the difference between the log BE/ME of the H portfolio and the log BE/ME of the L portfolio. The HML ROE is defined as the difference between the log ROE of the H portfolio and the log ROE of the L portfolio. These portfolios are constructed following Fama and French's (1993) methodology. The international HML portfolio is country balanced; that is, for each country, the stocks are first sorted into the six elementary portfolios based on country-specific breakpoints and all stocks in each elementary portfolio are value weighted. For each specification, we report three rows. The first row reports the OLS estimates of linear-regression coefficients of HML return on predictor variables. The OLS R^2 in the first row is adjusted for degrees of freedom. The second row reports the maximum-likelihood GLS estimates of the linear-regression coefficients of HML return on predictor variables. The GLS R^2 in the second row is the variance of fitted values (computed using GLS parameter estimates) divided by the sample variance of HML return. The third row shows the maximum-likelihood GLS coefficients in an exponential-linear conditional-variance model, shown in equation (19) of the text. *T*-statistics are in parentheses.

Panel A: U.S. HML return (1938–1997)

Parameter Estimate (t-Statistic)	Constant	Value Spread	Lagged ROE	Market BE/ME	Default Yield Spread	Value Spread* market BE/ME	R^2	St. dev. $E_{t-1}(R_t)$
OLS return coeff.	− 0.323 (− 2.4)	0.246 (2.8)					0.108	0.040
GLS return coeff.	− 0.376 (− 2.8)	0.287 (3.1)					0.160	0.047
GLS variance coeffi.	− 7.239 (− 4.2)	1.830 (1.6)						
OLS return coeff.	− 0.322 (− 2.4)	0.256 (2.8)	0.001 (0.2)				0.093	0.040
GLS return coeff.	− 0.360 (− 2.7)	0.292 (3.1)	0.003 (0.6)				0.156	0.045
GLS variance coeffi.	− 7.491 (− 4.4)	1.907 (1.6)	− 0.017 (− 0.3)					
OLS return coeff.	− 0.317 (− 2.4)	0.263 (3.0)				0.035 (1.1)	0.111	0.043
GLS return coeff.	− 0.414 (− 2.9)	0.334 (3.2)				0.044 (1.3)	0.226	0.054
GLS variance coeffi.	− 6.971 (− 4.0)	1.957 (1.7)				0.758 (1.8)		
OLS return coeff.	− 0.311 (− 2.4)	0.285 (3.0)	0.004 (0.7)			0.045 (1.3)	0.103	0.044
GLS return coeff.	− 0.425 (− 3.0)	0.376 (3.5)	0.006 (1.0)			0.056 (1.6)	0.254	0.058
GLS variance coeffi.	− 7.537 (− 4.4)	2.452 (2.0)	0.026 (0.4)			0.732 (1.6)		

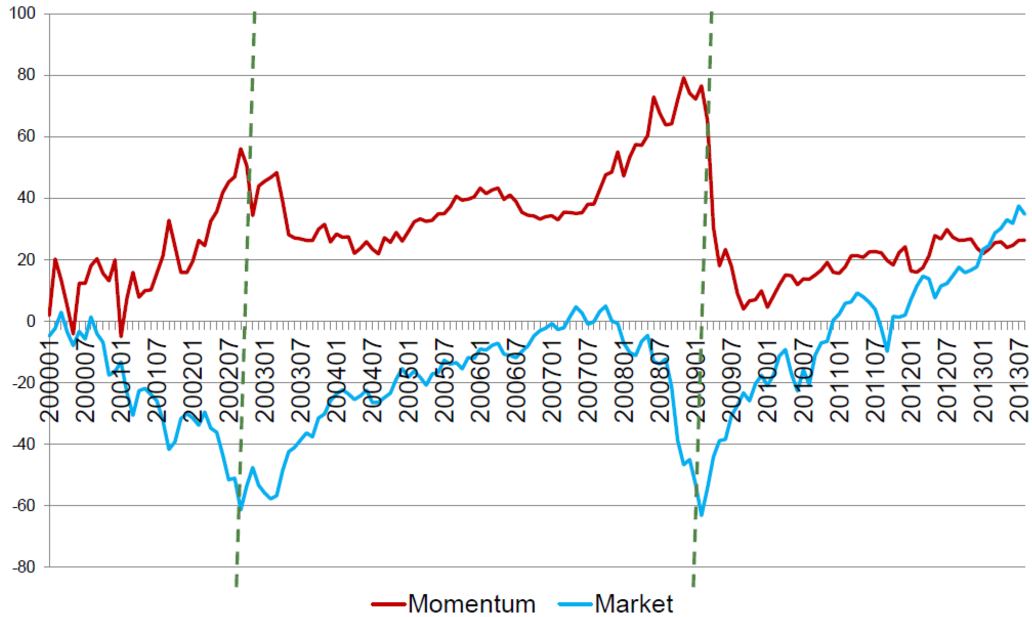
- The international evidence is consistent, but insignificant.

Panel B: Country-Balanced International HML Return (1983–1998)

Parameter estimate (t-statistic)	Constant	Value Spread	R^2	St. dev. $E_{t-1}(R_t)$
OLS return coeff.	− 0.117 (− 0.5)	0.131 (0.7)	− 0.030	0.013
GLS return coeff.	− 0.151 (− 0.8)	0.159 (1.0)	0.056	0.016
GLS variance coeffi.	− 8.438 (− 1.8)	2.309 (0.6)		

- Haddad, Kozak and Santosh (2020) extend this idea, and show that a broad range of long-short anomaly factors are predictable by their own lagged valuation spread.

- Momentum has very extreme downturns.



- Momentum crashes just when the market turns around after a downturn, see Daniel and Moskowitz (2016):

Rank	Month	WML_t	Mkt-2y	Mkt_t
1	1932-08	-0.7896	-0.6767	0.3660
2	1932-07	-0.6011	-0.7487	0.3375
3	2009-04	-0.4599	-0.4136	0.1106
4	1939-09	-0.4394	-0.2140	0.1596
5	1933-04	-0.4233	-0.5904	0.3837
6	2001-01	-0.4218	0.1139	0.0395
7	2009-03	-0.3962	-0.4539	0.0877
8	1938-06	-0.3314	-0.2744	0.2361
9	1931-06	-0.3009	-0.4775	0.1380
10	1933-05	-0.2839	-0.3714	0.2119
11	2009-08	-0.2484	-0.2719	0.0319

- Momentum (UMD) factor had -5.3% return in 2020.04 when the market recovered from the Covid-19 crash from 2020.02-2020.03.

- Market volatility is also a good predictor of momentum returns.

Table 5
Momentum Returns and Estimated Market Variance
Each column of this table presents the estimated coefficients and t-statistics for a time-series regression based on the following regression specification:

$$\tilde{R}_{\mathrm{WML},t} = \gamma_0 + \gamma_{\mathrm{B}} \cdot \mathrm{I}_{B,t-1} + \gamma_{\sigma^2_m} \cdot \hat{\sigma}^2_{m,t-1} + \gamma_{int} \cdot \mathrm{I}_{\mathrm{B,t-1}} \cdot \hat{\sigma}^2_{m,t-1} + \tilde{\epsilon}_t,$$

where $\mathrm{I}_{B,t-1}$ is the bear market indicator and $\hat{\sigma}^2_{m,t-1}$ is the variance of the daily returns on the market, measured over the 126-days preceding the start of month t. The regression is estimated using monthly data over the period 1927:07-2013:03. The coefficients $\hat{\gamma}_0$ and $\hat{\gamma}_B$ are $\times 100$ (i.e., are in percent per month).

	(1)	(2)	(3)	(4)	(5)
$\hat{\gamma}_0$	1.955	2.428	2.500	1.973	2.129
	(6.6)	(7.5)	(7.7)	(7.1)	(5.8)
$\hat{\gamma}_B$	-2.626		-1.281		0.023
	(-3.8)		(-1.6)		(0.0)
$\hat{\gamma}_{\sigma^2_m}$		-0.330	-0.275		-0.088
		(-5.1)	(-3.8)		(-0.8)
$\hat{\gamma}_{int}$				-0.397	-0.323
				(-5.7)	(-2.2)

- Moreira and Muir (2017) show that this result holds more broadly.
- If we scale anomaly returns by lagged volatility (measured based on daily returns over the last month), the Sharpe ratio increases significantly. Hence, average returns do not scale with volatility.
- Managed portfolios that take less risk when volatility is high produce larger alphas.

Table I. Volatility managed factor alphas. We run time-series regressions of each volatility managed factor on the non-managed factor $f_t^\sigma = \alpha + \beta f_t + \varepsilon_t$. The managed factor, f^σ, scales by the factors inverse realized variance in the preceding month $f_t^\sigma = \frac{c}{RV_{t-1}^2} f_t$. In Panel B, we include the Fama-French three factors as additional controls in this regression. The data is monthly and the sample is 1926-2015 for Mkt, SMB, HML, and Mom, 1963-2015 for RMW and CMA, 1967-2015 for ROE and IA, and 1983-2015 for the FX Carry factor. Standard errors are in parentheses and adjust for heteroscedasticity. All factors are annualized in percent per year by multiplying monthly factors by 12.

Panel A: Univariate regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Mkt^σ	SMB^σ	HML^σ	Mom^σ	RMW^σ	CMA^σ	FX^σ	ROE^σ	IA^σ
MktRF	0.61								
	(0.05)								
SMB		0.62							
		(0.08)							
HML			0.57						
			(0.07)						
Mom				0.47					
				(0.07)					
RMW					0.62				
					(0.08)				
CMA						0.68			
						(0.05)			
Carry							0.71		
							(0.08)		
ROE								0.63	
								(0.07)	
IA									0.68
									(0.05)
Alpha (α)	4.86	-0.58	1.97	12.51	2.44	0.38	2.78	5.48	1.55
	(1.56)	(0.91)	(1.02)	(1.71)	(0.83)	(0.67)	(1.49)	(0.97)	(0.67)
N	1,065	1,065	1,065	1,060	621	621	360	575	575
R^2	0.37	0.38	0.32	0.22	0.38	0.46	0.33	0.40	0.47
rmse	51.39	30.44	34.92	50.37	20.16	17.55	25.34	23.69	16.58

Panel B: Alphas also controlling for Fama-French 3 factors

Alpha (α)	5.45	-0.33	2.66	10.52	3.18	-0.01	2.54	5.76	1.14
	(1.56)	(0.89)	(1.02)	(1.60)	(0.83)	(0.68)	(1.65)	(0.97)	(0.69)

- This poses a challenge for several risk-based explanations because these portfolios take less risk in recessions (when volatility tends to be high).

- Stambaugh, Yu, and Yuan (2012) show that the performance of many anomalies is related to "sentiment" as defined by Baker and Wurgler (2006).

- Dynamics of sentiment

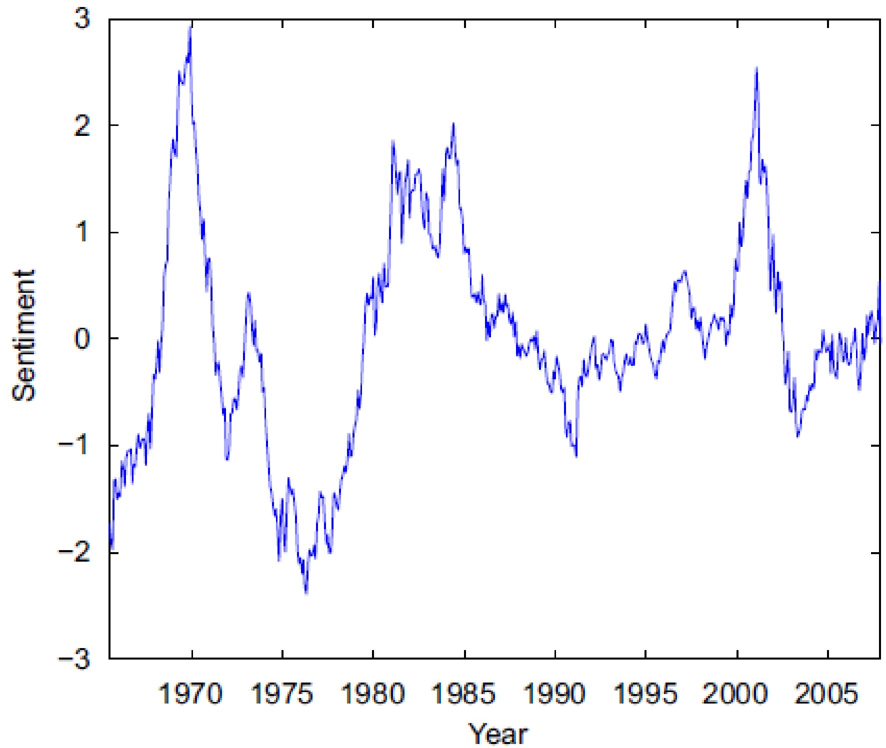


Fig. 1. The investor sentiment index from 1965:07 to 2007:12. The sentiment index is the first principal component of six measures: the closed-end fund discount, NYSE share turnover, the number of and the average of first-day returns on initial public offerings, the equity share in new issues, and the dividend premium. To control for macro-conditions, the six raw sentiment measures are regressed on the growth of industrial production, the growth of durable consumption, the growth of nondurable consumption, the growth of service consumption, the growth of employment, and a dummy variable for National Bureau of Economic Research recessions.

- Alphas (controlling for the market, size, and value) of anomaly strategies during periods of high and low sentiment on the long and the short end

$$R^e_{it} = \alpha_H d_{H,t-1} + \alpha_L d_{L,t-1} + bMKT + cSMB_t + dHML_t + \epsilon_{it}.$$

Anomaly	Long leg			Short leg			Long–short		
	High sentiment	Low sentiment	High –low	High sentiment	Low sentiment	High –low	High sentiment	Low sentiment	High –low
Failure probability	0.43	0.33	0.10	−1.65	−0.58	−1.07	2.08	0.91	1.17
	(2.52)	(2.33)	(0.44)	(−4.33)	(−1.81)	(−2.19)	(4.45)	(2.39)	(1.95)
Ohlson's O (distress)	0.25	0.16	0.09	−1.24	−0.60	−0.64	1.49	0.76	0.73
	(2.70)	(2.09)	(0.72)	(−5.29)	(−3.23)	(−2.16)	(6.13)	(3.77)	(2.32)
Net stock issues	0.28	0.11	0.17	−0.80	−0.12	−0.68	1.08	0.23	0.85
	(3.68)	(1.68)	(1.69)	(−4.86)	(−1.09)	(−3.42)	(6.19)	(1.79)	(3.90)
Composite equity issues	0.08	−0.03	0.11	−0.64	−0.17	−0.47	0.72	0.14	0.58
	(0.69)	(−0.31)	(0.72)	(−3.62)	(−1.57)	(−2.26)	(3.40)	(0.89)	(2.23)
Total accruals	0.19	0.34	−0.14	−0.70	0.02	−0.73	0.89	0.31	0.58
	(0.85)	(2.13)	(−0.53)	(−2.88)	(0.15)	(−2.53)	(3.02)	(1.33)	(1.60)
Net operating assets	0.22	0.27	−0.05	−0.87	−0.15	−0.72	1.09	0.42	0.67
	(1.36)	(2.04)	(−0.24)	(−4.94)	(−1.25)	(−3.40)	(4.78)	(2.20)	(2.30)
Momentum	0.66	0.60	0.06	−1.51	−0.76	−0.75	2.17	1.36	0.81
	(3.64)	(3.46)	(0.23)	(−4.03)	(−3.22)	(−1.69)	(4.46)	(3.87)	(1.35)
Gross profitability	0.46	0.41	0.05	−0.40	−0.06	−0.33	0.85	0.47	0.38
	(3.17)	(3.25)	(0.26)	(−2.43)	(−0.47)	(−1.59)	(3.77)	(2.23)	(1.24)
Asset growth	0.37	0.07	0.30	−0.82	−0.06	−0.76	1.18	0.13	1.05
	(2.23)	(0.38)	(1.29)	(−4.48)	(−0.48)	(−3.43)	(4.81)	(0.60)	(3.35)
Return on assets	0.49	0.27	0.23	−1.26	−0.51	−0.75	1.75	0.78	0.97
	(4.01)	(2.26)	(1.35)	(−3.98)	(−2.01)	(−1.88)	(5.00)	(2.66)	(2.16)
Investment-to-assets	0.01	0.32	−0.31	−0.73	−0.01	−0.72	0.74	0.33	0.41
	(0.09)	(2.53)	(−1.57)	(−4.31)	(−0.07)	(−3.34)	(3.75)	(1.76)	(1.54)
Combination	0.30	0.26	0.04	−0.92	−0.26	−0.66	1.22	0.52	0.70
	(5.62)	(5.40)	(0.62)	(−6.46)	(−2.95)	(−3.89)	(7.92)	(5.01)	(3.74)

- Anomaly strategies perform significantly better when investment sentiment is high than when it is low.

3. Estimating Factor Models

- See Cochrane chapters 11-14 for a detailed discussion on estimating factor models. We provide a high-level summary here.

- Consider the model

$$
\begin{aligned}
R^e_{i,t} &= a_i + \beta_i' F_t + \epsilon_{i,t}, & (2) \\
E(R^e_{i,t}) &= \alpha_i + \beta_i' \lambda. & (3)
\end{aligned}
$$

- The key prediction is that $\alpha_i = 0$, $\forall i$.

- We distinguish two cases

  1. The factors are excess returns themselves ("traded factors").

  2. The factors are non-traded, such as oil price shocks or consumption growth.

3.1. *Traded Factors*

- In the first case, we can estimate the model via time-series regressions of (2) and $\alpha_i = a_i$.

- We can use standard OLS results to test whether $\alpha_i = 0$.

- We are often more interested to test whether all pricing errors are jointly zero, $\alpha_1 = \cdots = \alpha_N = 0$.

- If we assume returns are i.i.d. normal, Gibbons, Ross, and Shanken (1989) show

$$\frac{T-N-K}{N}\left(1 + E_T(F_t)'\widehat{\Omega}^{-1}E_T(F_t)\right)^{-1}\widehat{\alpha}'\widehat{\Sigma}^{-1}\widehat{\alpha} \sim F_{N,T-N-K},$$

  where

  - N: Number of test assets.
  - K: Number of factors.
  - T: Length of the sample (time series).
  - Σ: Covariance matrix of the regression residuals.
  - $E_T(\cdot)$: Sample mean.
  - Ω: Covariance matrix of the factors.

- We can also use asymptotic theory, which does not require us to make assumptions about the finite-sample distribution of returns (i.e., normality) and leads to a χ^2_N test statistic,

$$T\left(1+E_T(F_t)'\widehat{\Omega}^{-1}E_T(F_t)\right)^{-1}\widehat{\alpha}'\widehat{\Sigma}^{-1}\widehat{\alpha}\sim\chi^2_N.$$

- Note that $Cov(\widehat{\alpha})^{-1}=T\left(1+E_T(F_t)'\widehat{\Omega}^{-1}E_T(F_t)\right)^{-1}\widehat{\Sigma}^{-1}$ (and analogously for the GRS statistic).

3.2. *Non-traded Factors*

- Recall the model

$$R^e_{i,t} = a_i + \beta_i' F_t + \epsilon_{i,t}, \quad (4)$$

$$E(R^e_{i,t}) = \alpha_i + \beta_i' \lambda, \quad (5)$$

where we want to test the null hypothesis that $\alpha_i = 0$.

- In case of traded factors, we have that $\lambda = E(F_t)$ and we can directly test whether the alphas, the intercepts, are jointly zero.

- In case of non-traded factors, it still holds under the null that the model prices all assets correctly

$$R^e_{it} = \beta_i' \lambda + \beta_i'(F_t - E(F_t)) + \epsilon_{it}.$$

- This implies that in the time-series regression,

$$R^e_{i,t} = a_i + \beta_i' F_t + \epsilon_{i,t},$$

we want to test the restriction

$$a_i = \beta_i'(\lambda - E(F_t)).$$

- However, the problem is that we do not have an estimate of λ and so we cannot directly test this restriction. In time-series regressions with traded factors, we have $\widehat{\lambda} = E_T(F_t)$.

- There are two common procedures in case of non-traded factors

    1. Cross-sectional (or, two-pass) regressions.

    2. Fama-MacBeth.

3.2.1. Cross-sectional regressions

- Cross-sectional regressions proceed in two steps:

    1. Estimate betas via time-series regressions as before, asset by asset. This delivers one full-sample $\widehat{\beta}_i$ for each asset.

    2. Estimate a cross-sectional regression of average returns on the test assets on the betas estimated from the first step (and a constant) to obtain the market price of risk estimate $\widehat{\lambda}$:
    $$E_T(R^e_{i,t}) = \alpha_i + \widehat{\beta}_i'\lambda.$$

- Form pricing errors for each test asset: $\alpha_i = E_T(R^e_{i,t}) - \widehat{\beta}_i'\lambda$

- We can now test whether the alphas from the second step are jointly zero via

$$\widehat{\alpha}' Cov(\widehat{\alpha})^{-1}\widehat{\alpha} \sim \chi^2_{N-K}.$$

We loose K degrees of freedom as we estimate K prices of risk, which mechanically ensures that K linear combinations hold exactly.

- In this test, we ignored that betas have been estimated in the first step. The estimated betas introduce an errors-in-variables problem. To address this, we can adjust the covariance matrix of the alphas using the Shanken (1992) correction (see Cochrane, Chapter 12).

$$Cov(\widehat{\alpha}) \;=\; \frac{1}{T}\left(I_N - \beta(\beta'\beta)^{-1}\beta'\right)\Sigma\left(I_N - \beta(\beta'\beta)^{-1}\beta'\right)(1+\lambda'\Sigma_f^{-1}\lambda),$$

  where the term $(1+\lambda'\Sigma_f^{-1}\lambda)$ is the modification of the standard OLS formulas due to the fact that betas are estimated in the first pass. In this notation, Σ_f is the covariance matrix of the factors and Σ the covariance matrix of the residuals.

- A similar correction can be derived for the estimated risk prices $\widehat{\lambda}$:

$$Cov(\widehat{\lambda}) \;=\; \frac{1}{T}\left((\beta'\beta)^{-1}\beta'\Sigma\beta(\beta'\beta)^{-1}(1+\lambda'\Sigma_f^{-1}\lambda)+\Sigma_f\right),$$

- Instead of using OLS, you can use GLS in the second step using $Cov(\alpha) = T^{-1}(\Sigma + \beta\Sigma_F\beta')$ as the error covariance matrix.

- Derivation:

  - The definition of α is given by $\alpha = \overline{R_t^e} - \beta\lambda$.

  - Use $\overline{R_t^e} = a + \beta\overline{F_t} + \overline{\epsilon_t}$ and, under the null of the model, $E(R_t^e) = \beta\lambda = a + \beta E(F_t)$.

  - The last two equations combined imply

$$\overline{R_t^e} = \beta\lambda + \beta(\overline{F_t} - E(F_t)) + \overline{\epsilon_t},$$

    implying $\alpha = \beta(\overline{F_t} - E(F_t)) + \overline{\epsilon_t}$.

  - Consequently, $Cov(\alpha) = T^{-1}(\Sigma + \beta\Sigma_F\beta')$.

- Why do we have $\overline{R_t^e}$ instead of $E(R_t^e)$? $\overline{R_t^e}$ is the dependent variable in the cross-sectional regression.

- Note that these procedures can also be used for traded factors, but it does not necessarily give you the same results.

- If you (i) include the traded factor as a test asset and (ii) use GLS in the second step, the estimates coincide.

3.2.2. Fama-MacBeth

- The second main method for testing factor models with non-traded factors is Fama-MacBeth regressions (Fama and MacBeth, 1973)

- There are three steps:

  1. Estimate the betas using time-series regressions. This can be done using rolling windows or full-sample time-series regressions.

  2. Run a cross-sectional regression **each period** of realized returns on the estimated betas from the first stage

  $$R_{i,t} = \alpha_{i,t} + \widehat{\beta}_{i,t}\lambda_t.$$

  3. Compute the time-series average of the risk prices and alphas to obtain the final estimates. The standard deviation of the means gives standard errors.

  $$\begin{aligned}\widehat{\alpha}_i &= T^{-1}\sum_t \alpha_{it}, \;\; \widehat{\lambda} = T^{-1}\sum_t \lambda_t, \\ Cov(\widehat{\alpha}) &= T^{-1}Cov_T(\alpha_t), \;\; Var(\widehat{\lambda}) = T^{-1}Cov_T(\lambda).\end{aligned}$$

  We can easily adjust the covariance matrices for serial correlation in the estimates if necessary.

- The test of the factor model then becomes

$$\widehat{\alpha}' Cov(\widehat{\alpha})^{-1}\widehat{\alpha} \sim \chi^2_{N-K}.$$

- Note: This ignores the estimation error in the betas.

- Fama-MacBeth regressions are also a common alternative to sorting stocks into portfolios. In that case, we are not only interested in testing the model but also in the slope estimates, $\widehat{\beta}_t$.

$$\widehat{\beta} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\beta}_t. \tag{6}$$

- To test whether a characteristic significantly predicts returns, we can compute the standard error of $\widehat{\beta}$ as:

$$Var(\widehat{\beta}) = \frac{1}{T}Var_T(\widehat{\beta}_t). \tag{7}$$

  Using the Newey-West estimator, you can account for persistence in slope (if there is any).

- Fama-MacBeth slopes are linear in returns,

$$\widehat{\beta}_t = (x_t'x_t)^{-1}x_t'R^e_{t+1} = w_t'R^e_{t+1},$$

  where w_t is a vector of portfolio weights.

- For a good example of using Fama-MacBeth regressions to understand the link between returns and characteristics, see Lewellen (2015).

- More generally, we can estimate asset pricing models using GMM.

- Cochrane Chapter 13-15 discusses connections between the various estimation approaches.

- GMM is attractive if we want to emphasize or de-emphasize certain moments. This is equivalent to weighted-least squares in cross-sectional regressions.

- This can be particularly important when we estimate a joint pricing model across asset classes where variances and risk premia differ significantly, for instance, across for bond and stock markets as in Koijen, Lustig, and Van Nieuwerburgh (2017).

- Recent JPE paper by Giglio and Xiu (2021) proposes a new method to test linear asset pricing models with non-traded factors that can deal with the (ever-present) concern of omitted factors. Proposes three-pass methodology that exploits:

  - the large dimensionality of available test assets

  - a rotation invariance result to correctly recover the risk premium of any observable factor, even when not all true risk factors are observed and included in the model

- There has been debate about whether to estimate factor models using portfolios or single stocks. The basic tradeoff is:
  - Sorting stocks into portfolios mitigates idiosyncratic noise and leads to more precise estimates of factor exposures, β_i.
  - Sorting stocks into portfolios reduces the spread in betas in the cross-section, which is what we need to estimate accurately estimate the risk prices, λ.
- Relevant references are Ang, Liu, and Schwarz (2020), Gagliardi, Ossola, and Scaillet (2016), and Cattanea, Crump, Farrell, Schaumburg (2016).
- Many of these papers use N, T asymptotics, using the argument that there are often more stocks than time periods.
- This is true, but the size distribution of firms is highly skewed (Gabaix, 2011), which means that the "value-weighted" or "effective" number of observations may be a lot smaller.

4. The Factor Zoo

- The factor literature has historically evolved as follows:

  1. CAPM.
     - Excess return on the market as the single factor.
  2. 3-factor Fama and French model (Fama and French, 1992).
     - Market factor + size (SMB) + value (HML).
  3. 4-factor Carhart model (Carhart, 1997).
     - Market factor + size (SMB) + value (HML) + momentum.
  4. 5-factor Fama and French model (Fama and French, 2015) or the related model by Hou, Xue, and Zhang (2015).
     - Market factor + size (SMB) + value (HML) + profitability + investment.

- Researchers sometimes add the Pastor and Stambaugh (2003) aggregate liquidity factor alongside the 3-, 4-, or 5-factor models.

- The additional factors are often designed to be roughly independent. For instance, Fama and French construct SMB and HML as

$$\begin{aligned} SMB &= \frac{1}{3}(SmallValue + SmallNeutral + SmallGrowth) \\ &\quad -\frac{1}{3}(BigValue + BigNeutral + BigGrowth), \\ HML &= \frac{1}{2}(SmallValue + BigValue) - \frac{1}{2}(SmallGrowth + BigGrowth), \end{aligned}$$

  where they first form six portfolios based on size (Small or Big) and book-to-market (Value, Neutral, Growth).

- Such factor models have also huge practical applications:

  – Many index funds and ETFs target these factors directly (”style investing” or “smart beta” strategies).

  – Provides a benchmark and style classification of mutual funds and hedge funds (see for instance the style box at Morningstar).

- Over the past 30 years, the anomalies literature has discovered many characteristics and/or factors s.t. when stocks are sorted into portfolios based on the characteristic or factor exposure, the long-short portfolio has an alpha w.r.t. the benchmark model

- The benchmark model has become more sophisticated over time: the CAPM, then the 3-factor, 4-factor, and 5-factor models.

- Major concern: There are too many factors. Cochrane's 2011 presidential address referred to this as the factor zoo.

- Pontiff and McLean (2016) study <u>97 variables</u> that (supposedly) predict returns in the cross-section.

Event	Market	Valuation	Fundamental
Change in Asset Turnover	52-Week High	Advertising/MV	Accruals
Change in Profit Margin	Age-Momentum	Analyst Value	Age
Change in Recommendation	Amihud's Measure	Book-to-Market	Asset Growth
Chg. Forecast + Accrual	Beta	Cash Flow/MV	Asset Turnover
Debt Issuance	Bid/Ask Spread	Dividends	Cash Flow Variance
Dividend Initiation	Coskewness	Earnings-to-Price	Earnings Consistency
Dividend Omission	Idiosyncratic Risk	Enterprise Component of B/P	Forecast Dispersion
Dividends	Industry Momentum	Enterprise Multiple	G Index
Down Forecast	Lagged Momentum	Leverage Component of B/P	Gross Profitability
Exchange Switch	Long-term Reversal	Marketing/MV	G-Score
Growth in Inventory	Max	Org. Capital	G-Score 2
Growth in LTNOA	Momentum	R&D/MV	Herfindahl
IPO	Momentum and Long-term Reversal	Sales/Price	Investment
IPO + Age	Momentum-Ratings		Leverage
IPO no R&D	Momentum-Reversal		M/B and Accruals
Mergers	Price		NOA
Post Earnings Drift	Seasonality		Operating Leverage
R&D Increases	Short Interest		O-Score
Ratings Downgrades	Short-term Reversal		Pension Funding
Repurchases	Size		Percent Operating Accrual
Revenue Surprises	Volume		Percent Total Accrual
SEOs	Volume Trend		Profit Margin
Share Issuance 1-Year	Volume Variance		Profitability
Share Issuance 5-Year	Volume-Momentum		ROE
Spinoffs	Volume/MV		Sales Growth
Sustainable Growth			Tax
Total External Finance			Z-Score
Up Forecast			
ΔCapex - ΔIndustry CAPEX			
ΔNoncurrent Op. Assets			
ΔSales - ΔInventory			
ΔSales - ΔSG&A			
ΔWork. Capital			

- They study the performance of these variables during two out-of-sample periods:

    1. Once the paper is publicly available, but before it is published.

    2. Following the publication of the article.

- Main regression:

$$R_{it} = \alpha_i + \beta_1 \text{Post Sample Dummy}_{it} + \beta_2 \text{Post Publication Dummy}_{it} + e_{it},$$

  where the Post Sample Dummy is 1 if the observation is after the original sample used in the paper but before the publication date. Post Publication Dummy is 1 if the observation is after the publication date in the journal.

- In-sample mean of the anomalies is 58.2bp per month.

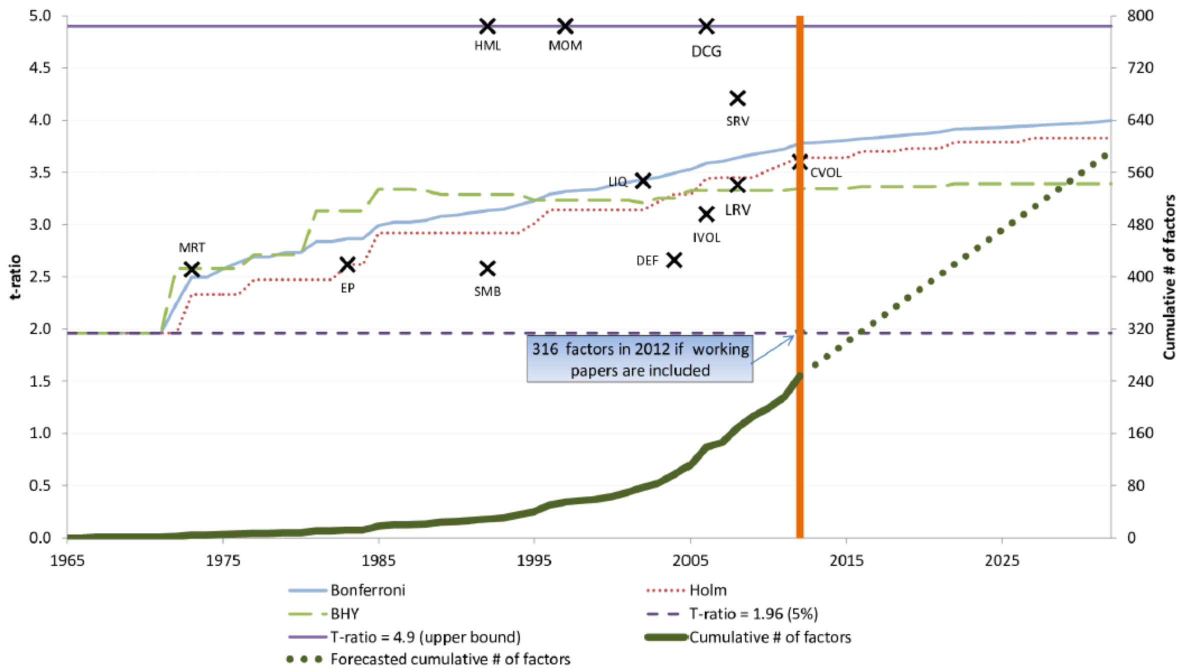
Variables	(1)	(2)
Post-Sample (S)	−0.150***	−0.180**
	(0.077)	(0.085)
Post-Publication (P)	−0.337***	−0.387***
	(0.090)	(0.097)

- Portfolio returns are 26% (=0.150/0.582) lower out-of-sample and 58% (=0.337/0.582) lower post-publication. Hence, they estimate a 32% lower return from publication-informed trading (58%-26%).

- The plethora of factor raises concerns about data mining. Statistical tests should account for "pre-testing," which is an old problem in econometrics. You cannot use conventional critical values to conclude that a a factor is priced or a variable predicts returns since that ignores *pre-testing*.

- Recently received significant attention via Harvey, Liu, and Zhu (2016) in cross-sectional asset pricing, the topic of Campbell Harvey's presidential address to the AFA.

- Number of factors and corrected $t-$statistics:



- Note that size, for instance, would have been insignificant at the moment of introduction if these critical values would have been used. Harvey, Liu, and Zhu (2016) suggest a $t-$statistic between 3.5 and 4.

- Of course, we do not know how many tests researchers ran before reporting the results, so we cannot correct perfectly for pre-testing.

- This is just for U.S. data. Suppose we assume for a second that the Fama-French 5-factor model does a good job pricing the cross-section of U.S. stock returns.

- Fama and French develop international versions of their model, see Fama and French (2015). They conclude:

  > Fama and French (2012) find that a Global version of the FF (1993) three-factor model does not explain international returns. A simple test produces the same negative conclusion for the five-factor model. We summarize the results, but do not present a table. We estimate 20 five-factor regressions in which the LHS returns are regional factors (five local factors for each of four regions) and the RHS returns are five Global factors. If the Global model describes expected returns, the 20 regression intercepts are indistinguishable from zero. In fact, five are more than three standard errors from zero, seven are more than two, and the GRS test of Gibbons, Ross, and Shanken (1989) says the probability the true intercepts are zero is tiny, zero to at least five decimal places.

- Hence, even the latest generation of factor models cannot price equity returns across four major regions (North America, Europe, Asia Pacific, and Japan).

- The main takeaway is that more structure and data are needed to discriminate among various factor models.

4.1. *Using holdings data to discriminate between models*

- One approach is to ask which factor model investors are using by using data on changes in stock holdings, or flows.
- Berk and van Binsbergen (2016) look at flows and returns together to study the risk model that mutual fund investors use.
- Basic idea:
  - There are decreasing returns to scale in active management and investors compete alphas to zero.
  - Investors learn about the manager's skill based on past performance, $R_{it} - R_{bt}$, where R_{it} is the manager's return and R_{bt} the risk-adjustment prescribed by an asset pricing model $\Rightarrow E_t([R_{i,t+1} - R_{b,t+1}]\alpha_{i,t+1}) > 0$.
  - Alphas are unobservable, but flows are. When the manager outperforms a model, the flow should be positive (and vice versa).
- Main conclusions:
  - The CAPM is the closest model to the model that investors use to decide on their capital allocation decisions.
  - The CAPM better explains flows than no model at all, indicating that investors do adjust for risk.
  - The CAPM also outperforms a naive model in which investors ignore beta and simply chase any outperformance relative to the market portfolio ($\beta = 1$ model).

- Authors' conclusion: The CAPM "is still the best method to use to compute the cost of capital of an investment opportunity" and should be used in making capital budgeting decisions.

- Paper by Barber, Huang, and Odean (2016) reaches essentially the same conclusion. They find that fund flows are more highly correlated with CAPM alphas than with other multi-factor alphas. They conclude that investors display limited sophistication.

- Agarwal, Green and Ren (2018) and Blocher and Molyboga (2017) carry out similar tests with samples of hedge funds.

- This research has recently attracted some controversy. Jagadeesh and Mangipudi (2021) argue that the strength of the predictive relation between fund flows and alphas is not a reliable moment to make inference about the true/best asset pricing model nor about investor sophistication.

  - Find that 4-factor alphas are most precise when the true betas are unknown to the econometrician, even if the true asset pricing model is the CAPM

4.2. *Using statistical techniques to "shrink" the factor zoo*

- A new and very active literature uses machine-learning tools to screen out the redundant and useless factors (due to data mining) from the truly useful asset pricing factors.

- Machine learning refers to collection of techniques:

  - High-dimensional models for statistical prediction
  - Regularization (penalization) methods for model selection and mitigation of overfit
  - Efficient algorithms for searching among vast number of potential model specifications, incl. non-linear relations

- More flexible than traditional econometric techniques, bringing hope of better approximating unknown and complex data generating process underlying equity risk premia, but at the risk of overfitting.

- See May 2020 special issue of RFS introduced by Karolyi and Van Nieuwerburgh (2020)

- Gu, Kelly, and Xiu (2020) compares machine learning techniques in asset pricing context
- Employs typical "big" data set
  - 30,000 individual stocks over 60 years from 1957-2016
  - 900 predictors
    - 94 characteristics
    - interactions of each characteristic with 8 aggregate time-series variables
    - 74 industry sector dummy variables
  - The predictors quickly increase in the thousands once (non-linear) interactions among them are considered.
- Generic model for excess returns:

$$R^e_{i,t+1} = E_t[R^e_{i,t+1}] + \epsilon_{i,t+1} = G(z_{i,t}) + \epsilon_{i,t+1}$$

  where $G(\cdot)$ is the same across time and across stocks, and only uses information at time t and for stock i.

4.2.1. Linear Model (with robust objective functions)

- Linear model: $G(z_{i,t};\theta) = z'_{i,t}\theta$.
- OLS minimizes the standard L_2-normed loss function:

$$\mathcal{L}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(r_{i,t+1} - G(z_{i,t};\theta)\right)^2$$

- Robust objective functions are an extension that allows to weight the different observations unequally, for example to deal with an unequal number of stocks across periods, or to use value weights (instead of equal weighting)
- Weighted least squares:

$$\mathcal{L}_W(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} w_{i,t} \left(r_{i,t+1} - G(z_{i,t}; \theta) \right)^2$$

- Objective function to deal with outliers/heavy tails of return distribution:

$$\mathcal{L}_H(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} H \left(r_{i,t+1} - G(z_{i,t}; \theta); \xi \right)$$

where the Huber loss function $H(\cdot)$ is the squared loss for small errors and the absolute loss for large errors, controlled by the tuning parameter ξ

$$H(x; \xi) = \begin{cases} x^2, & \text{if} \quad |x| \leq \xi \\ 2\xi|x| - \xi^2, & \text{if} \quad |x| > \xi \end{cases}$$

4.2.2. Penalized Linear Model

- The linear model is bound to fail with many predictors. When the number of predictors P approaches or exceeds the number of observations T, OLS becomes inefficient or even inconsistent. It overfits noise rather than extracting signal.

- Regularization: penalize the linear model to favor more parsimonious specifications:

$$\mathcal{L}(\theta) = \underbrace{\mathcal{L}(\theta)}_{Loss\ fcn} + \underbrace{\mathcal{P}(\theta;)}_{Penalty\ fcn}$$

- Elastic net penalty function with tuning parameters λ and ρ

$$\mathcal{P}(\theta; \lambda, \rho) = \lambda(1-\rho)\sum_{j=1}^{P}|\theta_j| + \frac{1}{2}\lambda\rho\sum_{j=1}^{P}\theta_j^2$$

- Special case 1: LASSO: $\rho = 0$, see Tibshirani (1996)
  - L_1 parameter penalization, on absolute values
  - **Selection method** that zeros out coefficients on subset of predictors

- Special case 2: Ridge regression: $\rho = 1$
  - L_2 parameter penalization, drawing all coefficient estimates closer to zero
  - **Shrinkage method** that prevents coefficients from becoming too large

4.2.3. Dimension reduction: PCR, PLS, RP-PCA

- Penalized linear models like the elastic net use shrinkage and variable selection to reduce dimensionality by forcing coefficients on most predictors to zero or near zero.
- This can produce suboptimal forecasts when predictors are highly correlated
- Forming *linear combinations of predictors* may be better to reduce noise and isolate the signal and to help de-correlate predictors
- Two classical dimension reducing techniques are Principal Components Regression (PCR) and Partial Least Squares (PLS)
- PCR:

$$w_j = \arg\max_w Var(Zw), \quad s.t. \quad w'w = 1, \quad Cov(Zw, Zw_l) = 0, \quad l = 1, 2, \cdots, j-1$$

Step 1 Find $K << P$ principal components of a set of predictors. Those are the K linear combinations of factors that explain most of the covariance of the predictors.

Step 2 Use the leading PCs in a linear predictive regression

- No guarantee that resulting factors are good return predictors.

- PLS, which goes back to Wold (1966), instead chooses K linear combinations of factors Z to have maximal predictive association with the forecast target, excess returns:

$$w_j = \arg\max_{w} Cov^2(R, Zw), \quad s.t. \quad w'w = 1, \quad Cov(Zw, Zw_l) = 0, \quad l = 1, 2, \cdots, j$$

- PLS sacrifices how well the K linear combinations of Z approximate Z to find components that are better return predictors.

- Risk-Premium-PCA: Lettau and Pelger (2020) consider another variant of PCA analysis, which they label RP-PCA

- The idea is to not only use information on the covariance of returns but also on the mean of returns (first+second moments)

- Let the $T \times N$ matrix of excess returns be R, the $T \times K$ matrix of principal components F and the $K \times N$ loadings of the assets on the factors Λ

$$R = F\Lambda' + e$$

- Rather than finding the PCs of the sample covariance matrix $\frac{1}{T}R'R - \bar{R}\bar{R}'$ by minimizing:

$$\min_{\Lambda,F} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(r_{i,t} - F_t\Lambda_i'\right)^2$$

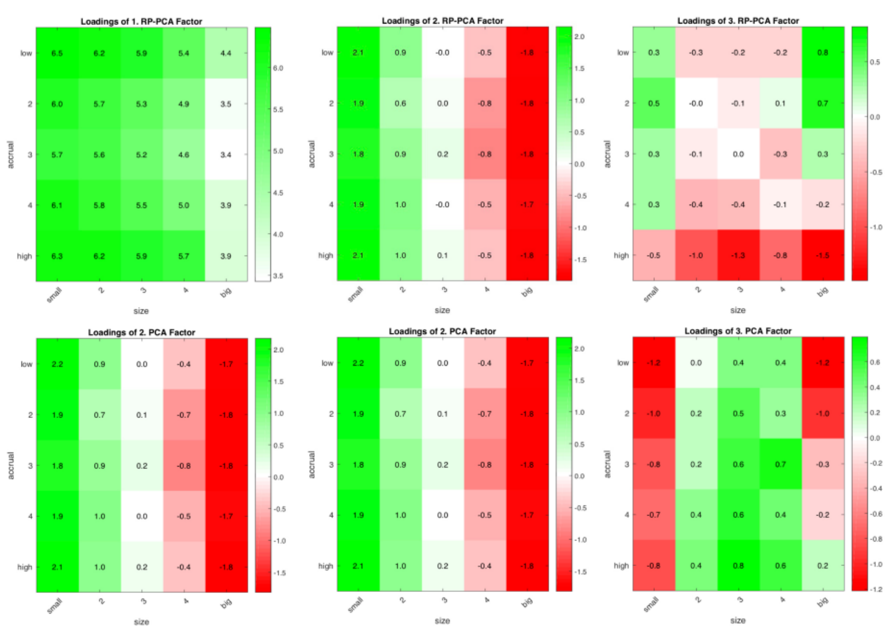
- RP-PCA penalizes the objective function for big pricing errors

$$\min_{\Lambda,F} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(r_{i,t} - F_t\Lambda_i'\right)^2 + \zeta \frac{1}{N} \sum_{i=1}^{N} \left(\overline{r}_i - \overline{F}\Lambda_i'\right)^2$$

- This is equivalent to applying PCA to the matrix $\frac{1}{T}R'R + \zeta\bar{R}\bar{R}'$

- Thus, if tuning parameter is $\zeta = -1$, back to standard PCA.
- Selects factors that not only have high Sharpe ratios (large PCs), but that can also explain the cross-section (small pricing errors). Standard PCA would throw out factors with smallish SRs but that are important for pricing the cross-section.
- Application considers 25 size/accruals double-sorted portfolios and 3 factors. PCA ($\zeta = -1$) has SR of 0.13 while RP-PCA ($\zeta = 10$) has SR of 0.24.
- Figure shows heat map of loadings on 3 factors. Loadings on first two factors are similar for PCA ($\zeta = -1$) and RP-PCA ($\zeta = 10$). Third PCA factor has no clear pattern and low SR (0.02), while third RP-PCA factor loads positively on low-accrual stocks and negatively on high-accrual stocks and high SR (0.20).

Figure 5: Heatmap of Factor Loadings for Size/Accrual Portfolios



Note: Size and Accrual: Heatmap of loadings. $K = 3$ statistical factors and RP- weight $\gamma = 10$.

4.2.4. Generalized Linear Model

- Linear models may poorly approximate potential non-linear relationship between predictors and returns.
- Non-parametric models of $G(z;\theta)$ can reduce this measurement error, at the risk of over-fitting and destabilizing the model out-of-sample
- Spline series expansion of predictors

$$G(z;\theta,p(\cdot)) = \sum_{j=1}^{P} p(z_j)'\theta_j$$

  where $p(\cdot) = (p_1(\cdot),\ldots,p_K(\cdot))$ is a $K \times 1$ vector of basis functions and θ is now $K \times P$

- Use least-squared objective function with Huber robustness modification and with group LASSO regularization

$$\mathcal{P}(\theta;\lambda,K) = \lambda \sum_{j=1}^{P} \left(\sum_{k=1}^{K} \theta_{j,k}^2 \right)^{1/2}$$

  Group LASSO selects either all K spline terms of a given predictor or none of them.

4.2.5. Boosted Regression Trees and Random Forests

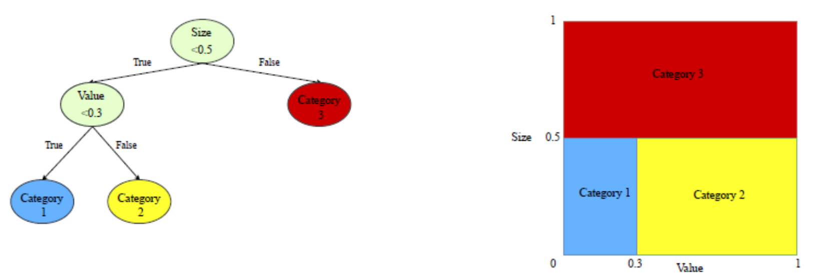
- Generalized linear model captures non-linearities of predictors but not interactions between predictors.
- Regression trees are popular way to incorporate multi-way predictor interactions.
- Tree with K leaves (terminal nodes) and depth L

$$G(z_{i,t};\theta,K,L) = \sum_{k=1}^{K} \theta_k 1_{z_{i,t}\in C_k(L)}$$

- To grow a tree is to find bins that best discriminate among the potential outcomes of the predictor.
- Simple example: tree based on size and BM. Tree first sorts observations based on size, then on BM.

$$G(z_{i,t};\theta,3,2) = \theta_1 1_{size_{i,t}<0.5} 1_{BM_{i,t}<0.3} + \theta_2 1_{size_{i,t}<0.5} 1_{BM_{i,t}\geq 0.3} + \theta_3 1_{size_{i,t}\geq 0.5}$$

Figure 1: Regression Tree Example



Note: This figure presents the diagrams of a regression tree (left) and its equivalent representation (right) in the space of two characteristics (size and value). The terminal nodes of the tree are colored in blue, yellow, and red, respectively. Based on their values of these two characteristics, the sample of individual stocks is divided into three categories.

- Simple objective function is to myopically minimize forecast error ("impurity") at the start of each branch C

$$H(\theta, C) = \frac{1}{|C|} \sum_{z_{i,t} \in C} (r_{i,t+1} - \theta)^2$$

  where $|C|$ is number of observations in set C

  Trees of depth L can capture $(L-1)$-way interactions

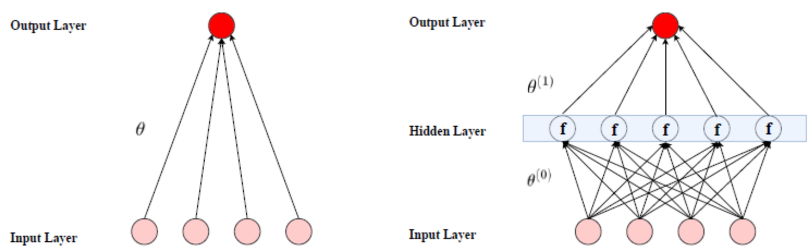
- Boosting is a first regularization method. It recursively combines forecasts from many shallow trees (small L). For example, gradient boosted regression trees (GBRT).

- Random forest is a second regularization method. Example of bootstrap aggregation or "bagging." Draw B bootstrap samples of data, fit regression tree to each, average their forecasts. Random forests reduce correlation among trees in different bootstrap samples by only considering a randomly drawn subset of predictors in each bootstrap sample.

4.2.6. Neural Networks

- Most powerful "deep learning" algorithms
- Also most complex and least transparent; prone to overfitting
- Combines an "input layer" of predictors with one or more "hidden layers" that interact and non-linearly transform the predictors, and "output layer" that aggregates hidden layers into ultimate outcome prediction.

Figure 2: Neural Networks



Note: This figure provides diagrams of two simple neural networks with (right) or without (left) a hidden layer. Pink circles denote the input layer and dark red circles denote the output layer. Each arrow is associated with a weight parameter. In the network with a hidden layer, a nonlinear activation function f transforms the inputs before passing them on to the output.

- At each "neuron," an "activation function" f transforms inputs into output

$$\begin{aligned} G(z;\theta) &= \theta_0^{(1)} + \sum_{k=1}^{5} x_k^{(1)} \theta_k^{(1)} \\ x_k^{(1)} &= f\left(\theta_{k,0}^{(0)} + \sum_{j=1}^{4} z_j^{(1)} \theta_{k,j}^{(0)}\right) \end{aligned}$$

Popular choice for f is rectified linear unit (RLU): $f(x) = x$ if $x > 0$ and zero otherwise.

4.2.7. Modified Fama-MacBeth

- The apparent good performance of several asset pricing models in the literature may be the spurious outcome of a weak identification problem. For more on the weak instruments problem applied to asset pricing, see Kleibergen, 2009.
- These weak or completely spurious factors may in turn make it more difficult to detect the true asset pricing factors.
- Bryzgalova (2016) replaces the second stage of Fama-MacBeth by a penalized version

$$\widehat{\lambda}_{penalized} = \arg\min_{\lambda} \frac{1}{2N} \left(E_T(R_t^e) - \widehat{\beta}\lambda \right)' W_T \left(E_T(R_t^e) - \widehat{\beta}\lambda \right) + \eta_T \sum_{j=1}^{K} w_j \mid \lambda_j \mid,$$

  - W_T is the weighting matrix and w_j is a measure of the correlation between portfolio returns and factor j.
  - For $\eta_T = 0$, we obtain the standard Fama-MacBeth estimator. For more severe penalties, we start shrinking risk premia to zero.
  - L_1 penalty on the λ's, essentially on the correlation of betas with average returns, a measure of factor strength.
  - Many macro factors tend to fall out because they have a weak correlation with stock returns.

				Fama-MacBeth estimator							Pen-FM estimator			
Model (1)	Factors (2)	p-value (Wald) (3)	GKR (2014) (4)	λ_j (5)	st.error (OLS) (6)	p-value (OLS) (7)	st.error (Shanken) (8)	p-value (Shanken) (9)	p-value (Bootstrap) (10)	R^2 (%) (11)	λ_j (12)	Shrinkage rate (Bootstrap) (13)	p-value (Bootstrap) (14)	R^2 (%) (15)
Panel A: tradable factors														
25 portfolios, sorted by size and book-to-market														
CAPM	Intercept	-	-	1.431***	0.4282	0.0008	0.4325	0.0009	0.002	19	1.4307***	0	0.002	19
	MKT	0	yes	-0.658	0.4256	0.1222	0.4764	0.1674	0.184		-0.658	0	0.184	
25 portfolios, sorted by size and book-to-market														
Fama and French (1992)	Intercept	-	-	1.252	0.2987	0	0.3054	0	0	70	1.2533	0	0	70
	MKT	0	yes	-0.703*	0.3035	0.0205	0.3721	0.0587	0.06		-0.704*	0	0.06	
	SMB	0	yes	0.145	0.0291	0	0.1424	0.3083	0.376		0.145	0	0.376	
	HML	0	no	0.43***	0.031	0	0.1376	0.0018	0.008		0.429***	0	0.008	
Panel B: nontradable factors														
25 portfolios sorted by size and book-to-market														
Lettau and Ludvigson (2001)	Intercept	-	-	3.24	0.6601	0	0.7261	0	0	16	3.8128	0.0000	0	0
	Δc	0.0061	no	0.216	0.1653	0.1922	0.1861	0.2468	0.398		0	0.889	0.975	
	Intercept	-	-	3.681**	0.9501	0.0001	1.4126	0.0092	0.012	31	4.106***	0	0.004	1
	cay	0.2006	no	-0.516	2.1813	0.813	3.3786	0.8786	0.518		0	0.958	0.976	
(scaled CAPM)	MKT	0	no	-0.056	0.9751	0.9545	1.6072	0.9724	0.72		-0.251	0	0.902	
	$cay\times$ MKT	0	no	1.136	0.3035	0.0002	0.4602	0.0136	0.398		0	0.839	0.851	
(scaled CCAPM)	Intercept	-	-	4.281***	0.7043	0	1.0093	0	0	70	4.958***	0	0	25
	cay	0.001	no	-0.125	0.2784	0.6526	0.41	0.7599	0.584		-0.385	0.319	0.513	
	Δc	0.7378	no	0.024	0.1095	0.8292	0.1619	0.884	0.654		0	0.689	0.721	
	$cay \times \Delta c$	0.0002	no	0.057	0.0171	0.0009	0.0254	0.0246	0.298		0	0.989	0.989	
(HC-CAPM)	Intercept	-	-	4.467***	0.9389	0	1.684	0.008	0.008	58	4.160***	0.0000	0	1
	MKT	0	no	-1.097	0.9322	0.2392	1.81	0.5445	0.798		-0.296	0	0.918	
	Δy	0.6566	no	1.259	0.3641	0.0005	0.6569	0.0552	0.1		0	0.815	0.815	
(scaled HC-CAPM)	Intercept	-	-	5.184***	0.9293	0	1.5628	0.0009	0.016	77	4.268***	0	0.002	7
	cay	0.0534	no	-0.445	0.2629	0.0908	0.4521	0.3254	0.502		0	0.987	0.987	
	MKT	0	no	-1.987	0.9226	0.0313	1.6995	0.2424	0.564		-0.438	0	0.8	
	Δy	0.3859	no	0.557	0.254	0.0282	0.4331	0.1982	0.23		0	0.792	0.792	
	$cay\times$ MKT	0	no	0.341	0.1841	0.0643	0.3225	0.2908	0.422		0	0.858	0.87	
	$cay \times \Delta y$	0.5135	no	-0.167	0.0678	0.014	0.1153	0.1485	0.422		-0.009	0.809	0.809	

4.2.8. Hyper-parameters

- All of these techniques have "hyper-parameters" or "tuning parameters," like λ and ρ in the elastic net approach, that need to be chosen. Some approaches such as neural networks (deep learning) have many of these parameters. How to choose them?
  - Little theoretical guidance.
  - Usual approach is to split data sample into three disjoint time periods.
  - Training sample to estimate the model for specific set of tuning parameters.
  - Validation sample to optimize over hyper-parameters by re-estimating model from training sample each time.
  - Results are then for a third, out-of-sample period.

4.2.9. Miscellaneous

- There is a host of new papers with related dimension reduction techniques which we will just mention in passing. This is an active literature.

  - Based on firm characteristics, Kelly, Pruitt, and Su (2019) find seven factors that are significant using their instrumented principal components analysis (IPCA).

  - Feng, Giglio, and Xiu (2020) find 14 out of 99 important factors with LASSO

  - Freyberger, Neuhierl, and Weber (2020) find significant improvements for pricing using a non-parametric LASSO method.

  - Han, He, Rapach, and Zhou (2018) introduce the use of combination forecasts and combination LASSO. Find that only 2 out of 94 characteristics matter for expected returns.

  - Using Bayesian LASSO, Kozak, Nagel, and Santosh (2020) find the best linear combinations of characteristics-based factors in a SDF framework and estimate the parameters by numerically solving a dual-penalty problem.

  - Chen, Pelger, and Zhu (2021) also employ an SDF framework, but use neural networks and extensive economic conditioning information (selected by an adversarial network).

  - He, Huang, Li, and Zhou (2021) use the reduced rank approach (RRA) to find the best 5-factor model to explain the cross-section of industry returns. The approach does not outperform the FF 5-factor model when it comes to pricing individual stock returns.

- Bandi, Chaudhuri, Lo, and Tamoni (2021) decompose systematic risk into its various spectral frequencies (horizons). The horizon of the risk exposures may provide an important dimension to reduce the dimensionality of pricing models. CAPM model based on low-frequency spectral betas works better than the FF 3-factor model.

- Lonn and Schotman (2018) form factor-mimicking portfolios (fmps) for macro-economic factors, essentially regressing them on 900 stock return portfolios. This is a small-T, large-N problem. They use a statistical learning technique called L_2-boosting, which uses an information criterion to penalize model complexity. They use these fmps to form a SDF in the spirit of Hansen and Jagannathan (1991).

- Ghosh, Julliard, Taylor (2017) non-parametrically estimate a time series for the SDF M_t extracted from the cross-section of excess stock returns R_t^e using a relative entropy minimization approach.

$$\arg\min_{M_t} E[M_t \log(M_t)] \quad s.t. \quad E[M_t R_t^e] = 0$$

The solution is:

$$\widehat{M_t} = \frac{\exp(\widehat{\theta}'_T R_t^e)}{\frac{1}{T}\sum_{t=1}^{T} \exp(\widehat{\theta}'_T R_t^e)}$$

where $\widehat{\theta}_T$ is the vector of Lagrange multipliers that minimize $\frac{1}{T}\sum_{t=1}^{T}\exp(\theta' R_t^e)$. The subscript T refers to the last period of the sample; the estimation can be done on a rolling window basis.

They then use this (single-factor) "I-SDF" out-of-sample for cross-sectional pricing and optimal asset allocation.

This works much better. The I-SDF has annual alphas of 14-16% relative to the FF5-factor model.

The SDF is particularly high *in those recessions in which the stock market does poorly*.

4.3. *Reflections on the Factor Zoo Literature*

1. Heading off model and method mining

   – Like the Harvey multiple testing critique, ML techniques require many implementation decisions. The fear is overfitting in the testing sample.

   – Providing Monte Carlo simulations, allowing for a larger set of hyper-parameters, broadening the sample of securities, and proving statistical properties of the estimator are all steps researchers can take to alleviate concerns.

2. Finding common ground across methods

   – Many methods suggest a modest number of factors

   – But different methods appear to result in different factors. We need a better understanding of why and how the factors that result from the various methods differ from each.

  - For example, isolating their common component, or focussing on correlation of a factor with a candidate SDF

3. The need for economic interpretability

   - The end goal should be that we arrive at an SDF that comes closer to achieving the mean-variance efficient portfolio and that the SDF be economically interpretable.
   - Economic interpretability remains a challenge for all ML approaches, and is particularly acute for the more black-box-like approaches such as neural networks.
   - Only with solid economic intuition can such a synthesis of empirical evidence serve as the basis for more realistic asset pricing theories.
   - A natural next step is to incorporate holdings data and to use ML to better understand what factors (and what information) drive individual and institutional portfolio choices. It is important to better understand how prices and expected returns reflect the information actual investors use to form their portfolios.