Factors or Fake?
A New Look at Anomalies

Siddhartha Chib[1], Shuhua Xiao[2], Lingxiao Zhao[3]

[1]Washington University in St. Louis
[2]City University of Hong Kong
[3]Peking University HSBC Business School

Any resemblance to real anomalies is purely coincidental . . . until we compute posterior probabilities.

Introduction

- The asset-pricing literature has produced hundreds of candidate return anomalies, indexed by $a_1, \ldots, a_n$.
- A growing body of work argues that *robust anomalies may be far fewer than previously believed*.
- Prominent contributions include Harvey, Liu, and Zhu (RFS 2016), Hou, Xue, and Zhang (RFS 2020), and Chordia, Goyal, and Saretto (RFS 2020).
- Our main is to revisit this screening question from a new perspective.

Existing Testing Paradigm

- Existing tests all follow the same recipe.
- Specifically, for each anomaly a_i, one sets up the hypotheses

$$H_{0,i}: \; a_i \text{ is spanned by a benchmark SDF,}$$

$$H_{1,i}: \; a_i \text{ is not spanned by the benchmark SDF,}$$

  where the benchmark is, for example, CAPM, FF3, or `FF6`.

- Rejection of $H_{0,i}$ is then interpreted as "discovery" (that a_i carries pricing information not captured by the benchmark SDF).

Existing Testing Paradigm (contd)

- It has been recognized, however, that one should only make claims for discovery after accounting for
  - multiple testing and false discoveries,
  - data snooping and publication bias,
  - balancing of Type I and Type II errors.
- For example, McLean and Pontiff (JF 2016), Linnainmaa and Roberts (RFS 2018), Giglio, Liao, and Xiu (RFS 2021), (Harvey and Liu, JF 2020; Chen, MS 2025).

Is There a Problem?

- Despite many statistical advances, there is a *conceptual* issue arising from how the hypotheses are formulated.
- In the current paradigm, "discovery" means that a_i is not spanned by the chosen benchmark.
- Such a claim can be justified only if the chosen benchmark is the true SDF.
- Why? Because an anomaly that is not spanned by one benchmark may be spanned by another benchmark that is closer to the true SDF.

Methodology

- We ask a different question: "can a_i be spanned by anything at all?" If a_i can be spanned by *any* admissible SDF, it cannot be an anomaly.

- This insight leads to a new pair of hypotheses:

$$H_{0,i}: \ a_i \text{ is not spanned by any admissible SDF,}$$
$$H_{1,i}: \ a_i \text{ is spanned by at least one admissible SDF.}$$

- This is a much tougher standard for an anomaly to survive.

- Importantly, evidence of spanning (the discovery) can be established without knowing the true SDF.

Frequentist Test? Not easily possible

- In the usual case, suppose we fix a candidate SDF model M.
- Spanning reduces to a standard test of pricing error:

$$H_0: \ \alpha = 0.$$

- The standard test statistic is t or F.
- Under H_0: central t or F distribution.
- Under the alternative: non-central distribution.
- For a *single* model, frequentist inference is well defined.

Our Null is Not a Single-Model Null

- Our null is

$$H_0: \; a_i \text{ is not spanned by any admissible model.}$$

- This means every admissible model has nonzero pricing error.
- A frequentist test would require:
  - testing spanning under every admissible model,
  - correcting for the model search.
- The global statistic becomes the minimum of many tests:

$$T^* = \min_j T_j,$$

where each T_j is non-central under H_0.

Why the Frequentist Route Breaks

- The null distribution of T^* depends on:
  - unknown noncentrality parameters,
  - dependence across models,
  - the size and geometry of the model space,
  - the search procedure itself.
- The test is therefore non-pivotal and not identifiable.
- There is no model-free sampling distribution.

Bayesian Resolution

- Bayes is natural here: it treats the hypotheses symmetrically and allows integration over the entire model space.
- We compute posterior probabilities of spanning by scanning all admissible SDF decompositions (Chib and Zeng, JBES 2020; Chib, Zeng and Zhao, JF 2020; Chib, Zhao and Zhou, MS 2024).
- We then apply Bayesian multiple testing with false discovery control (Müller, Parmigiani and Rice, 2007; Newton et al., 2004) to detect the *fake anomalies* among $\{a_i\}$.

- n candidate anomalies: $a_1, a_2, \ldots, a_n$
- For each anomaly a_i, we could simply compare the two models
  - FF6 unspanned, a_i spanned.

$$\texttt{FF6} = \boldsymbol{\lambda}^{(i)} + \boldsymbol{u}^{(i)},$$
$$a_i = \Gamma^{(i)}\, \boldsymbol{x}^{(i)} + \boldsymbol{\varepsilon}^{(i)}$$

  - a_i not in the span of FF6

$$\begin{pmatrix} \texttt{FF6} \\ a_i \end{pmatrix} = \boldsymbol{\lambda}^{(i)} + \boldsymbol{u}^{(i)}$$

- This comparison is incomplete, however. There are many distinct ways in which a_i can be spanned or not spanned by subsets of FF6.

- Define the augmented factor set

$$\boldsymbol{f}^{(i)} = (\texttt{FF6}, a_i)$$

- Consider all possible splits of $\boldsymbol{f}^{(i)}$ ($j = 1, 2, ..., J$, for $J = 127 = 2^7 - 1$):
  - factors that are unspanned $\boldsymbol{x}_j^{(i)}$
  - factors that are spanned $\boldsymbol{w}_j^{(i)}$
  - Each of these splits defines a model: $\mathbb{M}_j^{(i)} = (\boldsymbol{x}_j^{(i)}, \boldsymbol{w}_j^{(i)})$

Table 1: Candidate Splits for Factor Decomposition (Benchmark: FF6; one at a time).

Model	$\boldsymbol{x}$	$\boldsymbol{w}$
$\mathbb{M}_1^{(i)}$	Mkt	SMB, HML, RMW, CMA, MOM, a_i
$\mathbb{M}_2^{(i)}$	SMB	Mkt, HML, RMW, CMA, MOM, a_i
$\mathbb{M}_3^{(i)}$	HML	Mkt, SMB, RMW, CMA, MOM, a_i
$\mathbb{M}_4^{(i)}$	RMW	Mkt, SMB, HML, CMA, MOM, a_i
$\mathbb{M}_5^{(i)}$	CMA	Mkt, SMB, HML, RMW, MOM, a_i
$\mathbb{M}_6^{(i)}$	MOM	Mkt, SMB, HML, RMW, CMA, a_i
$\mathbb{M}_7^{(i)}$	a_i	Mkt, SMB, HML, RMW, CMA, MOM
$\mathbb{M}_8^{(i)}$	Mkt, SMB	HML, RMW, CMA, MOM, a_i
$\mathbb{M}_9^{(i)}$	Mkt, HML	SMB, RMW, CMA, MOM, a_i
...		
$\mathbb{M}_{13}^{(i)}$	Mkt, a_i	SMB, HML, RMW, CMA, MOM
...		
$\mathbb{M}_{120}^{(i)}$	Mkt, SMB, HML, RMW, CMA, MOM	a_i
...		
$\mathbb{M}_{127}^{(i)}$	Mkt, SMB, HML, RMW, CMA, MOM, a_i	$\varnothing$

Model-Level Spanning Decompositions

- For each anomaly a_i, each admissible decomposition $\mathbb{M}_j^{(i)} = (\boldsymbol{x}_j^{(i)}, \boldsymbol{w}_j^{(i)})$ corresponds to the system

$$\boldsymbol{x}_j^{(i)} = \boldsymbol{\lambda}_j^{(i)} + \boldsymbol{u}_j^{(i)}, \tag{1}$$

$$\boldsymbol{w}_j^{(i)} = \Gamma_j^{(i)} \boldsymbol{x}_j^{(i)} + \boldsymbol{\varepsilon}_j^{(i)}, \qquad j = 1, \ldots, 127. \tag{2}$$

- Each model encodes a distinct hypothesis about which factors (including a_i) are spanned and which are not.

- Using the Bayesian model scanning methods of Chib and Zeng (2020) and Chib, Zeng, and Zhao (2020), we compute the marginal likelihood and posterior probability of each model:

$$\Pr(\mathbb{M}_j^{(i)} \mid \text{data}) = \frac{m_j^{(i)}}{\sum_{\ell=1}^{J} m_\ell^{(i)}}, \qquad j = 1, \ldots, J, \; i = 1, \ldots, n.$$

- The posterior probabilities $\Pr(\mathbb{M}_j^{(i)} \mid \text{data})$ summarize evidence across *all* admissible spanning decompositions.
- For each anomaly a_i, we aggregate this information to obtain the posterior probability that a_i is spanned by *some* subset of `FF6`:

$$p_i \;=\; \Pr(a_i \text{ is spanned} \mid \text{data}) \;=\; \sum_{j:\, a_i \in \boldsymbol{w}_j^{(i)}} \Pr(\mathbb{M}_j^{(i)} \mid \text{data}), \qquad i = 1, \ldots, n.$$

- Thus, p_i measures the total posterior mass assigned to models in which a_i is spanned.

Bayesian EFDR-Based Selection

- The posterior probability p_i directly induces the *local false discovery rate* (lfdr):

$$\mathrm{lfdr}_i = \Pr(H_{0,i} \mid \text{data}) = 1 - p_i,$$

where $H_{0,i}$ denotes the hypothesis that a_i is *not* spanned.

- For any selected set $S \subset \{1, \ldots, n\}$ (interpreted as "spanned"), a false discovery occurs when $i \in S$ but $H_{0,i}$ is true.

- The posterior *expected false discovery rate (EFDR)* is

$$\mathrm{EFDR}(S \mid \text{data}) = \frac{\sum_{i \in S} \mathrm{lfdr}_i}{\max\{|S|, 1\}}.$$

- This criterion delivers a principled, multiplicity-aware selection of spanned anomalies.

- *EFDR: One-at-a-Time Selection*:
  - Sort $\mathrm{lfdr}_{(1)} \leq \cdots \leq \mathrm{lfdr}_{(n)}$.
  - For $k = 1, \ldots, n$, define

$$\mathrm{EFDR}(k) \;=\; \frac{1}{k}\sum_{\ell=1}^{k} \mathrm{lfdr}_{(\ell)} \;=\; \frac{1}{k}\sum_{\ell=1}^{k} \big(1 - p_{(\ell)}\big).$$

  - Choose

$$k^* \;=\; \max\{k :\; \mathrm{EFDR}(k) \leq q\},$$

    and set

$$\mathcal{S}_1^* \;=\; \{a_{(1)}, \ldots, a_{(k^*)}\}.$$

  - Exit the one-at-a-time testing with the unspanned set

$$\mathcal{A}_2^* = \{a_1, \ldots, a_n\} \;\setminus\; S_1^*$$

Proposition (One-at-a-Time Anomaly Tests)

For individual anomaly tests $\{a_i\}_{i=1}^{n}$*, a step-up selection rule based on posterior local false discovery rates (lfdr) satisfies the following:*

1. *It controls the expected false discovery rate (EFDR) of the selected set at level* q.
2. *Among all selection rules that satisfy this EFDR constraint, the step-up rule is Bayes-optimal in the sense of maximizing the expected number of true discoveries.*

Anomaly Spanning: Two at-a-Time Tests

- In practice, anomalies may be *jointly* spanned by the same SDF.
- We therefore extend the analysis to joint tests involving groups of anomalies.
- Let $\mathcal{A}_2^* \subset \{1, \ldots, n\}$ denote the set of anomalies remaining after the one-at-a-time screening stage.

Candidate Groups of Fixed Size s

- In the second step, fix a group size of $s = 2$.
- Consider all distinct groups of size s drawn from $\mathcal{A}_2^*$.
- Let

$$\mathcal{G}_1, \ldots, \mathcal{G}_{N_s}$$

  denote these groups, where

$$N_s = \binom{|\mathcal{A}_2^*|}{s}.$$

- Each $\mathcal{G}_i$ represents a candidate set of anomalies that may be *jointly* spanned.

Joint Spanning Hypotheses

- For each group $\mathcal{G}_i$, we test the joint spanning hypotheses:

$$H_{0,i}^{(s)}: \text{ none of the anomalies in } \mathcal{G}_i \text{ are spanned,}$$

$$H_{1,i}^{(s)}: \text{ all anomalies in } \mathcal{G}_i \text{ are jointly spanned.}$$

- Let

$$p_{i,s}^{\text{joint}} = \Pr\left(H_{1,i}^{(s)} \mid \text{data}\right)$$

  denote the posterior probability of joint spanning for group $\mathcal{G}_i$.

- The corresponding local false discovery rate is

$$\mathrm{lfdr}_{i,s} = 1 - p_{i,s}^{\text{joint}}.$$

- Pool the N_s groups $\{\mathcal{G}_i\}_{i=1}^{N_s}$ and sort their lfdrs:

$$\mathrm{lfdr}_{(1),s} \leq \mathrm{lfdr}_{(2),s} \leq \cdots \leq \mathrm{lfdr}_{(N_s),s}.$$

- For $k = 1, \ldots, N_s$, define the expected false discovery rate:

$$\mathrm{EFDR}_s(k) \;=\; \frac{1}{k} \sum_{\ell=1}^{k} \mathrm{lfdr}_{(\ell),s}.$$

- Selection rule: choose

$$k_s^* \;=\; \max\{k : \; \mathrm{EFDR}_s(k) \leq q\}, \qquad q \in (0,1).$$

- Given k_s^*, the set of anomalies classified as spanned at stage s is

$$\mathcal{S}_s^* = \bigcup_{\ell=1}^{k_s^*} \mathcal{G}_{(\ell)}.$$

- These anomalies are removed from further testing before proceeding to larger group sizes.

Proposition (Global EFDR Control)

Let $\mathcal{S}_1^*, \ldots, \mathcal{S}_S^*$ denote the subsets declared spanned at each stage, with increasing group sizes. Define the cumulative discovery set as

$$\mathcal{S}^* = \bigcup_{s=1}^{S} \mathcal{S}_s^*.$$

The posterior expected false discovery rate for the cumulative discovery set is

$$\mathrm{EFDR}\big(\mathcal{S}^* \mid \mathcal{D}\big) = \frac{1}{\max\big(|\mathcal{S}^*|, 1\big)} \sum_{\mathcal{S} \in \mathcal{S}^*} \mathrm{lfdr}_{\mathcal{S}}.$$

Under nested elimination, this global EFDR satisfies

$$\mathrm{EFDR}\big(\mathcal{S}^* \mid \mathcal{D}\big) \le q.$$

Empirical Analysis

"When you have eliminated the impossible, whatever remains, however improbable, must be the truth." — Sherlock Holmes

- We take the methodology to the data.
- The evidence shows that many anomalies are fake.
- The conclusions are robust across two benchmarks.
- An SDF built from the benchmark and the survivors improves both pricing fit and portfolio performance.

Data

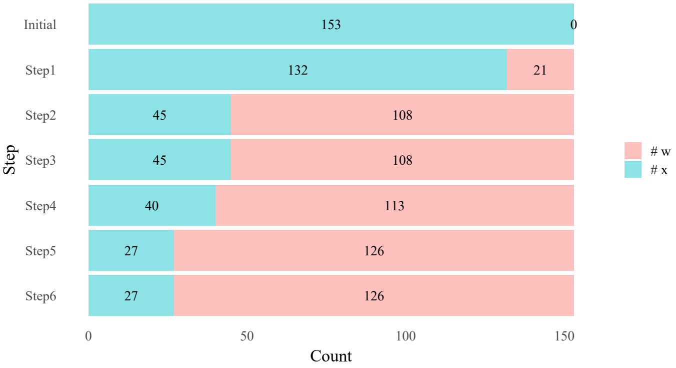
- 153 anomalies grouped into 13 thematic categories (Jensen, Kelly, and Pedersen, JF 2023)
- January 1985 through December 2024, $T = 480$ months.
- Training sample: the first 25% of observations ($T_{\text{tr}} = 120$)
- Test assets:
  - Fama-French 49 industry portfolios
  - 100 P-Tree portfolios (Cong, Feng, He and He, JFE 2025)
  - 120 ($2 \times 3 \times 20$) bivariate-sorted portfolios

First Benchmark: FF6

MKT	Market risk: Compensation for aggregate market exposure; CAPM backbone.
SMB	Size: Small minus big firms; captures size-related frictions and financing constraints.
HML	Value: High vs. low book-to-market firms; classic value (cheap vs. expensive).
RMW	Profitability: Robust minus weak operating performance; often interpreted as a quality factor.
CMA	Investment: Conservative vs. aggressive investment; grounded in q-theory of investment.
UMD	Momentum: Return continuation; winners minus losers.

Fake Anomalies: `FF6` Benchmark

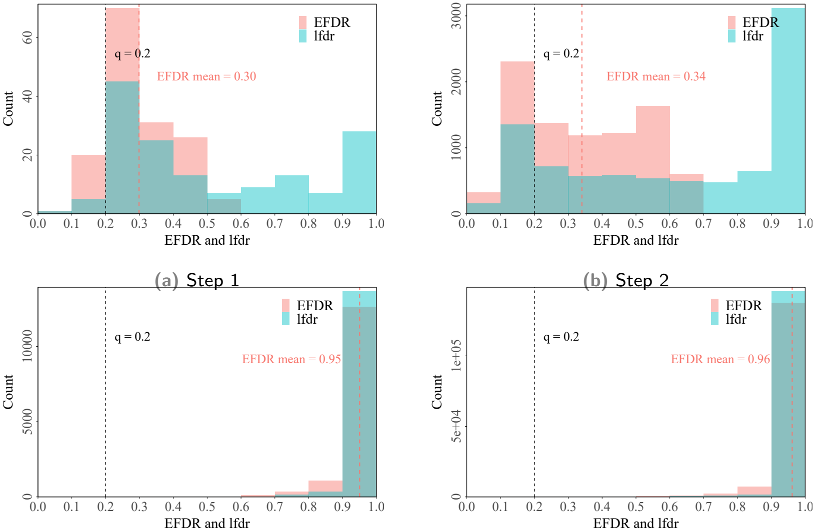
Figure 1: Counts of Fake Anomalies and Remaining Anomalies.



- 27 *last-stage anomalies* and 126 *fake anomalies*.
- Steps 1, 2, 4, and 5 contribute 21, 87, 5, and 13 new fake anomalies, respectively, while Step 3 yields none.

(ii) Fake Anomalies: `FF6` Benchmark

Figure 2: Histogram of EFDRs across Steps 1–4.



(a) Step 1

(b) Step 2

(c) Step 3

(d) Step 4

Why These?

Table 2: Number and Percentage of Anomalies.

Theme	$\#w$	$\#x$	$(\%)w$	$(\%)x$
Investment	22	0	100.0%	0.0%
Low risk	18	0	100.0%	0.0%
Momentum	8	0	100.0%	0.0%
Profit growth	11	1	91.7%	8.3%
Debt issuance	6	1	85.7%	14.3%
Short-term reversal	5	1	83.3%	16.7%
Size	4	1	80.0%	20.0%
Seasonality	10	2	83.3%	16.7%
Value	15	3	83.3%	16.7%
Accruals	3	3	50.0%	50.0%
Low leverage	7	4	63.6%	36.4%
Profitability	7	4	63.6%	36.4%
Quality	10	7	58.8%	41.2%
Total	126	27	82.4%	17.6%

- The closer a factor's economic content is to the benchmark, the more likely it is to be spanned and thus classified as $\boldsymbol{w}$.

(ii) Evaluating Fake Anomalies and Last-stage Anomalies

- The last-stage anomalies are either genuine factors or not: we cannot ascertain since the true SDF is unknown.
- We can assert: up to the level of our EFDR control, 126 anomalies can be spanned by `FF6` or by any proper subset of `FF6`

  ⇒ these can be classified as fake anomalies.

(ii) Evaluating Fake Anomalies and Last-stage Anomalies

- OOS predictive performance: predictive likelihood (Chib, Zeng and Zhao, MS 2024)
  - Sampling 1,000 instances of $5\boldsymbol{x}$ from $27\boldsymbol{x}$, and 1,000 instances of $5\boldsymbol{w}$ from remaining $\boldsymbol{w} \Rightarrow$ 1,000,000 pairs of $(5\boldsymbol{x}, 5\boldsymbol{w})$.
  - Panel A: in each pair, consider two models ((i) "True": risk factors are $5\boldsymbol{x}$; (ii) "Reverse":risk factors are $5\boldsymbol{w}$) and calculate the predictive likelihoods.
  - Panel B: Risk factors are $(5\boldsymbol{x}, \texttt{FF6})$("True"), or $(5\boldsymbol{w}, \texttt{FF6})$("Reverse").

Table 3: Statistics of Log Average of Predictive LLH for Factor Combinations

Factor Set		Mean	Std.	Median	2.5% qtile	97.5 % qtile
Panel A						
$(5\boldsymbol{x}, 5\boldsymbol{w})$	True	3241.5	113.5	3236.5	3033.8	3474.5
	Reverse	3236.7	113.8	3231.8	3028.1	3470.1
Panel B						
$(\texttt{FF6}, 5\boldsymbol{x}, 5\boldsymbol{w})$	True	5147.2	112.2	5143.3	4939.4	5375.0
	Reverse	5143.4	112.9	5139.7	4933.8	5372.3

(iii) Evaluating Fake Anomalies and Last-stage Anomalies

Table 4: Average Sharpe Ratios of Tangency Portfolios Across Different Factors.

$\boldsymbol{f}$	# Factor	# Comb	IS	OOS
Panel A. Only Anom.				
$5\boldsymbol{x}$	5	80,730	0.99	0.86
$5\boldsymbol{w}$	5	50,000	0.90	0.24
Panel B: Anom. & Bench. (Match with Method)				
$(5\boldsymbol{x}, \texttt{FF6})$	5+6	80,730	2.14	1.18
$(5\boldsymbol{w}, \texttt{FF6})$	5+6	50,000	1.93	0.69
Panel C: Anom. and Bench.				
$(27\boldsymbol{x}, \texttt{FF6})$	27+6	1	3.09	1.01
$(27\boldsymbol{w}, \texttt{FF6})$	27+6	50,000	3.06	0.48

(iv) Evaluating Fake Anomalies and Last-stage Anomalies

Table 5: Average Explanatory Power of Different Factors Across Test Assets.

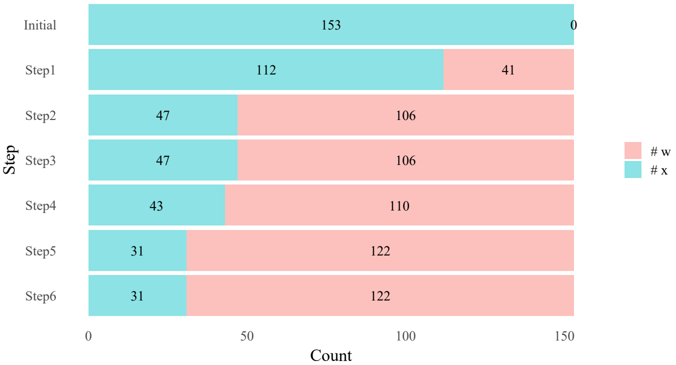
$\boldsymbol{f}$	# Factor	# Comb	Metrics				
			RMS α	RMSE	$\lvert\alpha\rvert$	GRS stat.	TotalR^2
Panel A. 49 Industry							
$(5\boldsymbol{x}, \texttt{FF6})$	11	80,730	0.0041	0.0474	0.0032	2.10	53.2
$(5\boldsymbol{w}, \texttt{FF6})$	11	50,000	0.0043	0.0480	0.0033	2.33	52.1
Panel B. 100 P-Tree							
$(5\boldsymbol{x}, \texttt{FF6})$	11	80,730	0.0038	0.0387	0.0028	2.93	65.8
$(5\boldsymbol{w}, \texttt{FF6})$	11	50,000	0.0038	0.0387	0.0028	3.07	65.8
Panel C. 120 Bisort							
$(5\boldsymbol{x}, \texttt{FF6})$	11	80,730	0.0019	0.0149	0.0015	13.46	95.0
$(5\boldsymbol{w}, \texttt{FF6})$	11	50,000	0.0019	0.0152	0.0015	14.67	94.8

Fake Anomalies: Another Benchmark

- NewB6: Retain Mkt but replace the other FF6 factors with:
  - the profitability (ROE) and investment (IA) factors from Hou, Xue, and Zhang (RFS, 2020) q-factor model
  - Pástor and Stambaugh (JPE, 2003) liquidity factor (LIQ)
  - Frazzini and Pedersen (JFE, 2014) betting-against-beta factor (BAB)
  - Asness and Frazzini (JPM, 2013) value factor (HMLD)

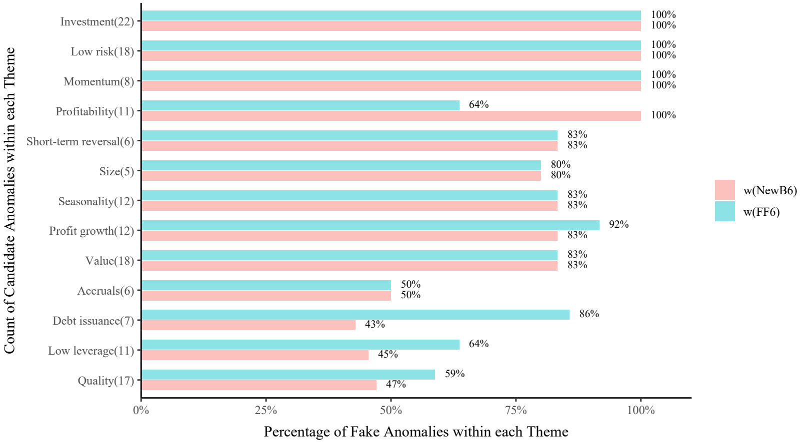
(ii) Fake Anomalies: Another Benchmark

Figure 4: Counts of Fake Anomalies and Remaining Anomalies (Benchmark: `NewB6`).



Fake Anomalies

Figure 5: Percentage of Fake Anomalies within each Theme.



- \# fake anomalies: 126(FF6) and 122(NewB6), respectively.
  The intersection of $\boldsymbol{w}_{\texttt{FF6}}$ and $\boldsymbol{w}_{\texttt{NewB6}}$ contains 116 elements.

$\Rightarrow$ Substantial overlap despite the two benchmarks sharing only Mkt.

What Remains?

Table 6: 21 Remaining Anomalies Across Benchmarks.

Names of *x*	**Theme**	**Description**
`cowc_gr1a`	Accruals	Change in current operating working capital
`oaccruals_at`	Accruals	Operating accruals
`oaccruals_ni`	Accruals	Percent operating accruals
`noa_at`	Debt issuance	Net operating assets
`cash_at`	Low leverage	Cash-to-assets
`netdebt_me`	Low leverage	Net debt-to-price
`rd_sale`	Low leverage	R&D-to-sales
`rd5_at`	Low leverage	R&D capital-to-book assets
`sale_emp_gr1`	Profit growth	Labor force efficiency
`cop_at`	Quality	Cash-based operating profits-to-book assets
`cop_atl1`	Quality	Cash-based operating profits-to-lagged book assets
`gp_at`	Quality	Gross profits-to-assets
`qmj`	Quality	Quality minus Junk: Composite
`qmj_safety`	Quality	Quality minus Junk: Safety
`sale_bev`	Quality	Assets turnover
`seas_6_10an`	Seasonality	Years 6-10 lagged returns, annual
`seas_11_15an`	Seasonality	Years 11-15 lagged returns, annual
`rd_me`	Size	R&D-to-market
`rmax5_rvol_21d`	Short-term reversal	Highest 5 days of return scaled by volatility
`fcf_me`	Value	Free cash flow-to-price
`div12m_me`	Value	Dividend yield

(ii) What Remains?

- Fundamental Quality:

  cash-based profitability (cop_at, cop_atl1), gross profitability (gp_at), composite quality (qmj, qmj_safety), asset turnover (sale_bev), labor productivity (sale_emp_gr1), accrual quality (cowc_gr1a, oaccruals_at, oaccruals_ni), and balance-sheet defensiveness (cash_at, netdebt_me)

- Management Decisions:

  net operating assets (noa_at); a suite of R&D intensity metrics (rd_sale, rd5_at, rd_me)

- Valuation Level:

  cash-flow yields: free-cash-flow yield (fcf_me) and dividend yield (div12m_me).

- Market Behavior and Patterns:

  short-horizon reversal (rmax5_rvol_21d), and long-lag seasonality (seas_6_10an, seas_11_15an).

Market prices of factor risks

$$\boldsymbol{x}^* = \underbrace{\boldsymbol{\lambda}_{x^*}}_{(6+21)\times 1} + \boldsymbol{\epsilon}_{x^*,t}, \quad \boldsymbol{\epsilon}_{x^*,t} \sim \mathcal{N}_{27}(\mathbf{0}, \boldsymbol{\Omega}_{x^*})$$

- Market prices of factor risks: $\boldsymbol{\Omega}_{x^*}^{-1}\boldsymbol{\lambda}_{x^*}$

Posterior summary of Market Prices of Factor Risks

Name	Theme	Post. mean	Post. std.	Post. median	2.5% qtile	97.5% qtile
Mkt		6.5	2.1	6.5	**2.4**	**10.7**
SMB		4.9	3.1	4.8	-1.3	11.1
HML		-0.1	4.1	-0.1	-8.2	8.0
RMW		18.4	5.5	18.3	**7.8**	**29.3**
CMA		16.5	5.1	16.5	**6.8**	**26.6**
MOM		4.8	2.0	4.8	**0.9**	**8.7**
cowc_gr1a	Accruals	1.5	5.0	1.5	-8.3	11.4
oaccruals_at	Accruals	-6.5	7.0	-6.5	-20.2	7.1
oaccruals_ni	Accruals	8.7	6.7	8.6	-4.3	21.8
noa_at	Debt issuance	3.6	5.2	3.6	-6.5	13.8
cash_at	Low leverage	5.8	6.3	5.8	-6.6	18.3
netdebt_me	Low leverage	5.9	6.2	5.9	-6.2	17.9
rd_sale	Low leverage	5.9	5.3	5.9	-4.5	16.3
rd5_at	Low leverage	-10.8	5.5	-10.7	-21.5	-0.2
sale_emp_gr1	Profit growth	-5.4	3.9	-5.4	-13.2	2.2
cop_at	Quality	-0.3	14.2	-0.3	-28.2	27.4
cop_atl1	Quality	18.0	14.4	18.0	-10.1	46.3
gp_at	Quality	-23.7	8.2	-23.7	-40.0	-7.7
qmj	Quality	8.6	6.2	8.6	-3.5	21.0
qmj_safety	Quality	-1.7	5.2	-1.7	-12.0	8.6
sale_bev	Quality	11.0	6.8	11.0	-2.2	24.3
seas_6_10an	Seasonality	8.6	2.6	8.6	**3.5**	**13.8**
seas_11_15an	Seasonality	7.2	3.1	7.1	**1.1**	**13.4**
rd_me	Size	8.7	3.3	8.7	**2.3**	**15.3**
rmax5_rvol_21d	Short-term reversal	6.7	2.4	6.7	**2.0**	**11.6**
fcf_me	Value	4.8	4.0	4.8	-2.9	12.7
div12m_me	Value	-1.3	5.4	-1.3	-11.8	9.3

Summary

- Determining whether a return signal is a *true anomaly* requires knowledge of the true SDF.
- As a result, proving *non-spanning* is, in principle, impossible.
- In contrast, proving spanning only requires exhibiting an admissible factor set that prices the signal.
- Our analysis shows that, rather than an anomaly zoo, we are left with a small, well-behaved petting farm.

Summary: way to do it

- Burden shifting:

$$H_{0,i}: \; a_i \notin \boldsymbol{w} \;\; \text{(not spanned)}, \qquad H_{1,i}: \; a_i \in \boldsymbol{w} \;\; \text{(spanned)}.$$

- Model uncertainty:
  - Integrate over all spanned/unspanned partitions of factors.
  - Evidence is summarized by posterior model probabilities, not by a single benchmark test.
- Multiplicity:
  - EFDR-controlled selection of spanned sets.

Summary

- FF6 stepwise scan:
  - 126 anomalies classified as spanned.
  - 27 anomalies remain unspanned.
- Investment relevance:
  - FF6 augmented with 5–6 late-stage survivors delivers higher in-sample and out-of-sample tangency Sharpe ratios.
  - Cross-sectional pricing fit dominates that of size-matched fake anomaly sets.

Summary

- Cross-benchmark robustness:
  - Union of spanned sets across `FF6` and `NewB6`: 132 anomalies.
  - Residual unspanned set: 21 anomalies.
- SDF implications:
  - Eight factors (four from `FF6`, four from the survivor set) have strictly positive MPR credible intervals.