PHBS WORKING PAPER SERIES

Factors or Fake? A New Look at Anomalies and the Replication Crisis

Siddhartha Chib
Washington University

Shuhua Xiao
City University of Hong Kong

Lingxiao Zhao
Peking University

November 2025

Working Paper 20251102

Abstract

Based on the insight that declaring an anomaly not spanned presumes knowledge of the true SDF, we reverse conventional testing: discovery now corresponds to returns being spanned by a given factor set. We develop a Bayesian procedure that, for single anomalies and sequentially for pairs, triplets, and larger sets, computes posterior spanning probabilities and selects discoveries while controlling the expected false discovery rate (EFDR) at each stage. Applied to the U.S. dataset of Jensen et al. (2023) comprising 153 anomalies, our method classifies 132 as spanned. The evidence tempers claims of new factors and highlights redundancy and masking in the anomaly zoo.

Keywords: Bayesian multiple testing, expected false discovery rate control, model uncertainty, posterior spanning probabilities, spanned anomalies
JEL Classification: C11, C12, G10, G12

Peking University HSBC Business School
University Town, Nanshan District
Shenzhen 518055, China

Factors or Fake? A New Look at Anomalies and the Replication Crisis*

Siddhartha Chib[a,*], Shuhua Xiao[b], Lingxiao Zhao[c]

[a]*Olin School of Business, Washington University in St. Louis, 1 Bookings Drive, St. Louis, MO 63130.*

[b]*College of Business, City University of Hong Kong. Tat Chee Avenue, Kowloon 999077, Hong Kong.*

[c]*Peking University HSBC Business School, University Town, Nanshan District, Shenzhen 518055, P.R.China.*

Abstract

Based on the insight that declaring an anomaly not spanned presumes knowledge of the true SDF, we reverse conventional testing: discovery now corresponds to returns being spanned by a given factor set. We develop a Bayesian procedure that, for single anomalies and sequentially for pairs, triplets, and larger sets, computes posterior spanning probabilities and selects discoveries while controlling the expected false discovery rate (EFDR) at each stage. Applied to the U.S. dataset of Jensen et al. (2023) comprising 153 anomalies, our method classifies 132 as spanned. The evidence tempers claims of new factors and highlights redundancy and masking in the anomaly zoo.

Keywords: Bayesian multiple testing, Expected false discovery rate control, Model uncertainty, Posterior spanning probabilities, Spanned anomalies

JEL: C11, C12, G10, G12

*We thank the seminar and conference participants at the 4th Hong Kong Conference for Fintech, AI, and Big Data in Business, for their valuable comments. Zhao L.'s research is supported by the Shenzhen Outstanding Talent Program for Scientific and Technological Innovation – Doctoral Basic Research Startup Grant (RCBS20231211090520033).

*Corresponding author

Email addresses: chib@wustl.edu (Siddhartha Chib), shxiao3-c@my.cityu.edu.hk (Shuhua Xiao), lingxiao@phbs.pku.edu.cn (Lingxiao Zhao)

Factors or Fake? A New Look at Anomalies and the Replication Crisis

Abstract

Based on the insight that declaring an anomaly not spanned presumes knowledge of the true SDF, we reverse conventional testing: discovery now corresponds to returns being spanned by a given factor set. We develop a Bayesian procedure that, for single anomalies and sequentially for pairs, triplets, and larger sets, computes posterior spanning probabilities and selects discoveries while controlling the expected false discovery rate (EFDR) at each stage. Applied to the U.S. dataset of Jensen et al. (2023) comprising 153 anomalies, our method classifies 132 as spanned. The evidence tempers claims of new factors and highlights redundancy and masking in the anomaly zoo.

Keywords: Bayesian multiple testing, Expected false discovery rate control, Model uncertainty, Posterior spanning probabilities, Spanned anomalies

JEL: C11, C12, G10, G12

1. Introduction

Over the past two decades, the empirical asset pricing literature has seen a proliferation of reported return anomalies, cross-sectional patterns in average returns that appear unexplained by standard factor models. This explosion of findings has sparked concerns about data mining, and whether such findings can be replicated, raising fundamental questions: Which anomalies are genuine and robust? Which merely reflect statistical noise or sample-specific overfitting?

In this paper, we introduce a new approach to anomaly evaluation, to distinguish between spanned and unspanned anomalies in a principled and scalable way. The key insight driving our approach is that it is statistically easier to verify that an anomaly is spanned rather than to argue that it is unspanned. Spanning means that the anomaly can be explained by the factor set, and this can be checked directly, even if the factor set is not optimal or complete. In contrast, even if an anomaly is unspanned by a given set of factors, it is possible that it can be spanned by another, more optimal set of risk factors. Thus, asserting that an anomaly is unspanned is inherently more challenging. Therefore, our framework reverses the usual logic. Specifically, the null is that the anomaly cannot be spanned, and discovery corresponds to the finding that it *can* be spanned.

To implement our approach in high dimensions and under model uncertainty, we develop a Bayesian sequential expected false discovery rate (EFDR) procedure. In this procedure, we compute posterior spanning probabilities for single anomalies, then sequentially for pairs, triplets, and larger sets, using all admissible partitions in the manner of Chib and Zeng (2020), that split the benchmark factors and anomalies into *spanned* and *unspanned* components. EFDR control is applied within a stepwise screening loop in which at each round, we remove factors with high posterior spanning probability (fake anomalies) and retain those that remain unspanned across rounds. This approach results in sharper, benchmark-robust inference with formal decision-theoretic guarantees about which anomalies are spanned and which are not.

Literature. One strand of the recent work emphasizes publication bias and data-snooping

in anomaly detection (e.g., Chen, 2025), which can be viewed as addressing uncertainty about the anomaly space itself. The dominant approach for anomaly detection is to determine whether an anomaly delivers incremental pricing power.[1] This is approached by setting the null that alpha is zero (the anomaly is spanned) and using t-tests for statistical significance. Several concerns about this approach have been noted in the literature. The frequentist test can only fail to reject the null or reject the null. Thus, in this framework, it is not possible to ever claim that the anomaly is spanned and, without a coherent alternative, it is not possible to know what is discovered when the null is rejected. The statistical power of these tests can be low, with a further loss under multiple testing, as discussed in Harvey, Liu, and Zhu (2016), Harvey and Liu (2020), and Harvey et al. (2020). This leads to fragile inferences (Lewellen et al., 2010; Harvey, 2017).

A second key issue with existing approaches is model uncertainty arising from incomplete knowledge of the stochastic discount factor (SDF). Because the true SDF is unknown, the benchmark factors used to evaluate anomalies are typically ad hoc and rarely justified.[2] Benchmark choice strongly affects which anomalies appear to survive (Fama and French, 2016) and which funds seem to exhibit skill (Lehmann and Modest, 1987), yet most studies only report results under a small number of alternative benchmarks as robustness checks (Yan and Zheng, 2017; Chordia et al., 2020).

Contributions. Recognizing that the conventional hypothesis-testing architecture and pervasive benchmark uncertainty make it nearly impossible to establish that an anomaly truly earns a significant alpha, we take the opposite perspective and reverse the hypotheses: the null is that the anomaly is *not spanned* by the benchmark factors, and the alternative is that it *is spanned*. Beyond the novelty of this idea, this inversion is natural for two reasons. First, declaring an anomaly to be genuine presumes knowledge of the true SDF, an object that is essentially unattainable in practice. In contrast, spanning, unlike alpha significance,

[1] Common diagnostics include time-series alpha tests, cross-sectional Fama–MacBeth regressions that estimate characteristic prices, and direct tests of high-minus-low (H–L) portfolio returns.

[2] Two benchmarks dominate applied work: the capital asset pricing model (CAPM; Sharpe (1964); Lintner (1965)) and the Fama–French three- and five-factor models (`FF3`, `FF5`; (Fama and French, 1993, 2015)). Variants include the Carhart four-factor model (Carhart, 1997) and the q-factor model (Hou et al., 2015).

is a directly testable restriction in the Bayesian framework and does not depend on the optimality of the benchmark SDF. Second, when confronted with a collection of putative anomalies, it is more coherent to begin from the null that they are anomalies and to seek evidence of spanning, rather than to assume the opposite. Applying our method to the U.S. dataset of Jensen et al. (2023), which contains 153 candidate anomaly factors, we find that 132 can be spanned, evidence that is consistent with a broader replication crisis in empirical asset pricing.

The rest of the paper is organized as follows. Section 2 introduces our Bayesian spanning framework, beginning with single anomalies and extending sequentially to pairs, triplets, and larger sets under EFDR control. Section 3 deals with the empirical analysis and shows that most of the anomalies given in Jensen et al. (2023) are spanned, leaving a much smaller residual set. Section 4 evaluates economic performance via predictive likelihoods and out-of-sample portfolio tests and reports robustness to alternative benchmarks. Section 5 concludes with implications for replication and the identification of genuine risk factors.

2. Methodology: Which Anomalies Can Be Spanned?

Our methodology for determining which anomalies can be spanned makes extensive use of the Bayesian model comparison framework developed in Chib and Zeng (2020) and Chib, Zeng, and Zhao (2020). This framework is used to sequentially evaluate whether anomalies are priced, that is, spanned by a given factor model. We begin by testing each anomaly individually alongside `FF6`, computing the posterior probability of spanning for each case. We then apply Bayesian EFDR control to select anomalies with strong evidence of being spanned. These are excluded from further testing. We next examine all remaining pairs of anomalies in combination with `FF6`, applying the same logic again. We continue this process iteratively, with the candidate set shrinking and the factor set expanding at each step, until no further anomalies are identified as spanned. Thus, up to the level of our EFDR control, these are the anomalies that may be classified as fake.

The residual set that remains may contain genuine factors or non-factors, but this

distinction cannot be known with certainty given that the true SDF is unobserved. Moreover, it is possible that another starting benchmark set of factors could span some of these remaining anomalies. To evaluate this possibility, we repeat the pruning process below using an alternative benchmark factor set and take the union of the spanned anomalies across benchmarks. Because our Bayesian EFDR-based approach imposes a strict burden of proof for declaring an anomaly spanned, the resulting spanned sets are expected to be largely stable across benchmark choices.

2.1. Anomaly Spanning: One at a Time

Suppose we have n candidate anomalies $a_1, a_2, \ldots, a_n$. For each anomaly a_i, we consider the augmented factor set

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

We now consider all possible splits of $\boldsymbol{f}^{(i)}$ into factors that are unspanned $\boldsymbol{x}_j^{(i)}$ and factors that are spanned $\boldsymbol{w}_j^{(i)}$, where $j = 1, 2, ..., J$, for $J = 127 = 2^7 - 1$, denotes the jth such split. Each of these splits defines a model that we indicate by $\mathbb{M}_j^{(i)}$. It is important to bear in mind that in some of these splits, a_i is an element of $\boldsymbol{x}_j^{(i)}$, and in other splits it is an element of $\boldsymbol{w}_j^{(i)}$.

In Table A.1 we enumerate all 127 possible splits of $\boldsymbol{f}^{(i)} = (\texttt{FF6}, a_i)$. The *unshaded rows* correspond to models in which a_i is included in the unspanned set $\boldsymbol{x}_j^{(i)}$; these are denoted by $\mathcal{M}_0 = \{\mathbb{M}_7^{(i)}, \mathbb{M}_{13}^{(i)}, \mathbb{M}_{18}^{(i)}, \ldots, \mathbb{M}_{126}^{(i)}, \mathbb{M}_{127}^{(i)}\}$. In contrast, the *shaded rows* correspond to models where a_i appears in the spanned set $\boldsymbol{w}_j^{(i)}$, denoted by $\mathcal{M}_1 = \{\mathbb{M}_1^{(i)}, \mathbb{M}_2^{(i)}, \mathbb{M}_3^{(i)}, \ldots, \mathbb{M}_{114}^{(i)}, \mathbb{M}_{120}^{(i)}\}$.

For each given anomaly a_i, the typical model $\mathbb{M}_j^{(i)}$is given by

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

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

where there is no intercept in the second model since $\boldsymbol{w}_j^{(i)}$ by definition is spanned by $\boldsymbol{x}_j^{(i)}$.

It is useful to note that included among these $J = 127$ models are the two conventional benchmarks typically compared in a Bayesian test of whether a_i is priced by FF6. The first is model $\mathbb{M}_{120}^{(i)}$ in Table A.1. This specification places all six FF6 factors in the unspanned set and a_i in the spanned set:

$$\underbrace{\texttt{FF6}}_{=\boldsymbol{x}_{120}^{(i)}} = \boldsymbol{\lambda}_{120}^{(i)} + \boldsymbol{u}_{120}^{(i)}, \tag{3}$$

$$\underbrace{a_i}_{=\boldsymbol{w}_{120}^{(i)}} = \Gamma_{120}^{(i)}\, \boldsymbol{x}_{120}^{(i)} + \boldsymbol{\varepsilon}_{120}^{(i)}, \tag{4}$$

so that a_i lies in the span of FF6. The other is model $\mathbb{M}_{127}^{(i)}$ in Table A.1. This specification treats a_i as an additional (unspanned) factor alongside FF6:

$$\underbrace{\begin{pmatrix} \texttt{FF6} \\ a_i \end{pmatrix}}_{=\boldsymbol{x}_{127}^{(i)}} = \boldsymbol{\lambda}_{127}^{(i)} + \boldsymbol{u}_{127}^{(i)}, \qquad \boldsymbol{w}_{127}^{(i)} = \varnothing. \tag{5}$$

in which case a_i is not in the span of FF6. A conventional Bayesian analysis would compare $\mathbb{M}_{120}^{(i)}$ and $\mathbb{M}_{127}^{(i)}$. Our approach is more comprehensive. It goes beyond this two-model comparison by embedding both within the larger collection $\{\mathbb{M}_j^{(i)} : j = 1, 2, \ldots, 127\}$, i.e., all possible splits of $\boldsymbol{f}^{(i)} = (\texttt{FF6}, a_i)$ into unspanned and spanned components.

Under Gaussianity of the errors and the priors given in Chib and Zeng (2020) and Chib, Zeng, and Zhao (2020), Bayesian fitting of each of the J models is straightforward. Importantly, from this Bayesian fitting, the marginal likelihood of the data under model $\mathbb{M}_j^{(i)}$, denoted as $m_j^{(i)}$, is available in closed form. Assuming a discrete uniform prior over the J models, given by $\Pr(\mathbb{M}_j^{(i)}) = 1/J$, the posterior probability of each model is available as

$$\Pr(\mathbb{M}_j^{(i)} \mid \text{data}) = \frac{m_j^{(i)}}{\sum_{\ell=1}^{J} m_\ell^{(i)}}, \qquad \text{for } j = 1, \ldots, J, \text{ and } i = 1, \ldots, n. \tag{6}$$

Given these posterior probabilities, we compute the total posterior probability that

anomaly a_i is spanned by any subset of the FF6 factors as follows:

$$\begin{aligned} p_i &= \Pr(a_i \text{ is spanned} \mid \text{data}) \\ &= \sum_{j:a_i \in \boldsymbol{w}_j^{(i)}} \Pr(\mathbb{M}_j^{(i)} \mid \text{data}) \ , \ i = 1, 2, \ldots, n, \end{aligned} \tag{7}$$

where the summation is taken over all models in which $a_i \in \boldsymbol{w}_j^{(i)}$. A large value of p_i indicates strong evidence that anomaly a_i is priced (i.e., spanned).

Since there are m anomaly-specific tests, it is crucial to address multiplicity to avoid a high rate of false discoveries. We adopt the expected false discovery rate (EFDR) procedure from the Bayesian multiple testing literature (e.g., Müller et al., 2007; Newton et al., 2004), which delivers a data-adaptive selection threshold and an explicit bound on the expected fraction of mistakes among the declared discoveries.

For each anomaly $i = 1, \ldots, n$, we test

$$H_{0,i} : \ a_i \notin \boldsymbol{w} \ \ (\text{not spanned}), \qquad H_{1,i} : \ a_i \in \boldsymbol{w} \ \ (\text{spanned}). \tag{8}$$

The roles of null and alternative are reversed relative to common frequentist treatments: our discovery is *spanning* (pricing), which frequentist tests cannot support directly. Moreover, as seen in Table A.1, $|\mathcal{M}_0^{(i)}| = 64$ and $|\mathcal{M}_1^{(i)}| = 63$; under a uniform prior over the 127 models, the prior probabilities of $H_{0,i}$ and $H_{1,i}$ are nearly equal, so posterior evidence is not dominated by prior asymmetry.

Let $p_i = \Pr(H_{1,i} \mid \text{data})$ be the posterior probability that anomaly i is spanned, given the data. Define the *local false discovery rate* (lfdr) as

$$\text{lfdr}_i \ = \ \Pr(H_{0,i} \mid \text{data}) \ = \ 1 - p_i. \tag{9}$$

If a rule selects a set $S \subset \{1, \ldots, n\}$ as "spanned," a false discovery occurs when $i \in S$ but $H_{0,i}$ is true; the posterior probability of that mistake is lfdr_i. Hence, the posterior expected

number of false discoveries is $\sum_{i \in S} \mathrm{lfdr}_i$, and the posterior expected false discovery rate is

$$\mathrm{EFDR}(S \mid \mathcal{D}) \;=\; \frac{\sum_{i \in S} \mathrm{lfdr}_i}{\max\{|S|, 1\}}. \tag{10}$$

Declaring "spanned" whenever $p_i \geq \tau$ ignores multiplicity and correlation across candidates. With many tests, some p_i will be large by chance, and a fixed τ provides no control over the expected fraction of mistakes among selections. In contrast, EFDR explicitly targets the decision problem "maximize discoveries subject to $\mathrm{EFDR} \leq q$," and sets the cutoff adaptively from the joint evidence.

The EFDR selector is formulated as follows. Sort the lfdr_i in increasing order:

$$\mathrm{lfdr}_{(1)} \leq \mathrm{lfdr}_{(2)} \leq \cdots \leq \mathrm{lfdr}_{(n)}, \tag{11}$$

and for $k = 1, \ldots, n$ define the running EFDR

$$\mathrm{EFDR}(k) \;=\; \frac{1}{k} \sum_{\ell=1}^{k} \mathrm{lfdr}_{(\ell)} \;=\; \frac{1}{k} \sum_{\ell=1}^{k} \big(1 - p_{(\ell)}\big). \tag{12}$$

Choose the largest k^* with $\mathrm{EFDR}(k^*) \leq q$ for a prespecified level $q \in (0, 1)$ (e.g., $q = 0.2$), and classify $a_{(1)}, \ldots, a_{(k^*)}$ as spanned. This controls the posterior EFDR at level q and yields a data-adaptive, multiplicity-aware selection.

We formally summarize this procedure, and its decision-theoretic optimality, in the following proposition.

Proposition 1 (Bayes–optimal EFDR–controlled selection)**.** *Let*

$$k^* = \max\left\{k \in \{0, \ldots, n\} : \; \frac{1}{k} \sum_{\ell=1}^{k} \mathrm{lfdr}_{(\ell)} \leq q\right\}$$

(with the empty sum for $k = 0$ taken as 0), and set $S^* = \{(1), \ldots, (k^*)\}$. Then:

1. *(*EFDR control*)* $\mathrm{EFDR}(S^* \mid \textit{data}) \leq q$.

2. *(*Bayes optimality*) Among all S with* $\mathrm{EFDR}(S \mid \textit{data}) \leq q$, *$S^*$ maximizes the posterior*

expected number of true discoveries,

$$\sum_{i \in S} p_i = \sum_{i \in S} \big(1 - \mathrm{lfdr}_i\big).$$

Proof. (1) By construction, $\mathrm{EFDR}(S^* \mid \text{data})$ equals the average of the k^* smallest lfdr values, which is $\leq q$. (2) Fix $k = |S|$. If S omits some $\mathrm{lfdr}_{(j)}$ while including a larger lfdr, swapping in $\mathrm{lfdr}_{(j)}$ weakly tightens the constraint and strictly increases $\sum_{i \in S}(1 - \mathrm{lfdr}_i)$. Thus the optimal S of size k is $\{(1), \ldots, (k)\}$; feasibility requires $\mathrm{EFDR}(k) \leq q$. Choosing the largest feasible k yields S^*. □

2.2. Anomaly Spanning: Two at a Time

In the second step of our procedure, we consider the spanning of two anomalies at a time. This step is important because stopping after singleton tests can miss two economically common situations. First, *masking*: two highly correlated anomalies may each look only weakly spanned in isolation, yet be *jointly* spanned once evaluated together. Second, *redundancy*: multiple proxies for the same latent risk can yield inflated marginal evidence. Joint testing de-duplicates such signals by requiring them to be priced *together*. A pairwise spanning stage therefore improves power against masked alternatives, reduces false discoveries from redundancy, and preserves multiplicity control by applying EFDR at the pair level.

Let $b_1, \ldots, b_m$, where $m = (n - k^*)$, denote the anomalies that were not classified as spanned in Step 1. We now consider all $n^* = \binom{m}{2}$ distinct pairs (b_i, b_k) with $i < k$. For each pair, we construct the augmented factor set

$$\boldsymbol{f}^{(i,k)} = (\mathtt{FF6}, b_i, b_k)$$

For each factor set $\boldsymbol{f}^{(i,k)}$, there are $J^* = 2^8 - 1 = 255$ possible nontrivial partitions into spanned factors $\boldsymbol{w}_j^{(i,k)}$ and unspanned factors $\boldsymbol{x}_j^{(i,k)}$. However, we now restrict attention only to models in which:

- **either** both b_i and b_k are in $\boldsymbol{x}_j^{(i,k)}$ (jointly unspanned).
- **or** both b_i and b_k are in $\boldsymbol{w}_j^{(i,k)}$ (jointly spanned),

We *exclude* from consideration all models in which one of b_i, b_k is spanned and the other is not. This focuses the inference on joint pricing behavior and removes ambiguity from mixed-spanning configurations.

Each model $\mathbb{M}_j^{(i,k)}$ in this restricted class takes the form

$$\boldsymbol{x}_j^{(i,k)} = \boldsymbol{\lambda}_j^{(i,k)} + \boldsymbol{u}_j^{(i,k)}, \tag{13}$$

$$\boldsymbol{w}_j^{(i,k)} = \Gamma_j^{(i,k)} \boldsymbol{x}_j^{(i,k)} + \boldsymbol{\varepsilon}_j^{(i,k)}, \tag{14}$$

with $\boldsymbol{w}_j^{(i,k)}$ spanned by construction, and no intercept in the second equation.

It is useful to emphasize that included among these $J_{\text{joint}} = 127$ admissible models (see Table A.2) are the following two models.

The first is model $\mathbb{M}_{120}^{(i,k)}$ in Table A.2. This specification places all six `FF6` factors in the unspanned set and both anomalies in the spanned set:

$$\underbrace{\texttt{FF6}}_{=\boldsymbol{x}_{120}^{(i,k)}} = \boldsymbol{\lambda}_{120}^{(i,k)} + \boldsymbol{u}_{120}^{(i,k)}, \tag{15}$$

$$\underbrace{\begin{pmatrix} b_i \\ b_k \end{pmatrix}}_{=\boldsymbol{w}_{120}^{(i,k)}} = \Gamma_{120}^{(i,k)} \boldsymbol{x}_{120}^{(i,k)} + \boldsymbol{\varepsilon}_{120}^{(i,k)}, \tag{16}$$

so that the pair (b_i, b_k) lies in the span of `FF6` (jointly spanned).

The second is model $\mathbb{M}_{127}^{(i,k)}$ in Table A.2. This specification treats both anomalies as

additional (unspanned) factors alongside FF6:

$$\underbrace{\begin{pmatrix} \texttt{FF6} \\ b_i \\ b_k \end{pmatrix}}_{=\boldsymbol{x}_{127}^{(i,k)}} = \boldsymbol{\lambda}_{127}^{(i,k)} + \boldsymbol{u}_{127}^{(i,k)}, \qquad \boldsymbol{w}_{127}^{(i,k)} = \varnothing, \tag{17}$$

in which case (b_i, b_k) is not in the span of FF6 (jointly unspanned).

We could just compare only $\mathbb{M}_{120}^{(i,k)}$ and $\mathbb{M}_{127}^{(i,k)}$ to find the probability of spanning. Our approach is more comprehensive: we embed both within the larger collection $\{\mathbb{M}_j^{(i,k)} : j = 1, \ldots, 127\}$, i.e., all admissible splits of $\boldsymbol{f}^{(i,k)} = (\texttt{FF6}, b_i, b_k)$ into unspanned and spanned components and find the probability of spanning from these splits, as we now explain.

Let $m_j^{(i,k)}$ denote the marginal likelihood of model $\mathbb{M}_j^{(i,k)}$, computable using the closed-form expressions described in Chib (1995) and applied in Chib and Zeng (2020). Under a uniform prior over the restricted set of models (now smaller than 255), we compute the posterior probability of model $\mathbb{M}_j^{(i,k)}$ as

$$\Pr(\mathbb{M}_j^{(i,k)} \mid \text{data}) = \frac{m_j^{(i,k)}}{\sum_l m_l^{(i,k)}}, \tag{18}$$

where the sum is over all models in the restricted space for the pair (b_i, b_k).

For each pair (b_i, b_k), define the joint inclusion probability:

$$p_{i,k}^{\text{joint}} = \sum_{j:\{b_i,b_k\}\subset \boldsymbol{w}_j^{(i,k)}} \Pr(\mathbb{M}_j^{(i,k)} \mid \text{data}), \tag{19}$$

which is the posterior probability that both b_i and b_k are jointly spanned under models in the restricted space.

We now frame the hypothesis test for the pair as follows:

- Null hypothesis $H_0^{(i,k)}$: *neither* b_i nor b_k is spanned.

- Alternative hypothesis $H_1^{(i,k)}$: *both* b_i and b_k are spanned.

Table A.2 presents the restricted set of model splits in which both (b_i, b_k) are either included in $\boldsymbol{x}_j^{(i,k)}$ or in $\boldsymbol{w}_j^{(i,k)}$. The unshaded rows correspond to the null hypothesis, while the shaded rows represent the alternative, comprising 64 and 63 models, respectively. Under a uniform prior over all models, the prior probabilities assigned to the null and alternative hypotheses are approximately equal, consistent with the one-at-a-time testing framework.

The local false discovery rate for the pair is defined as:

$$\text{lfdr}_{i,k}^{\text{joint}} = 1 - p_{i,k}^{\text{joint}}. \tag{20}$$

Let $n^* = \binom{m}{2}$ be the number of distinct anomaly pairs. We sort the local fdrs in increasing order:

$$\text{lfdr}_{(1)}^{\text{joint}} \leq \text{lfdr}_{(2)}^{\text{joint}} \leq \cdots \leq \text{lfdr}_{(n^*)}^{\text{joint}}. \tag{21}$$

Then for each $k = 1, \ldots, n^*$, compute the expected false discovery rate:

$$\text{EFDR}(k) = \frac{1}{k} \sum_{\ell=1}^{k} \text{lfdr}_{(\ell)}^{\text{joint}}. \tag{22}$$

Let k^{**} be the largest integer such that $\text{EFDR}(k^{**}) \leq q$, for a pre-specified threshold $q \in (0, 1)$ (e.g., $q = 0.2$).

The top k^{**} pairs are declared as jointly spanned:

$$\mathcal{P}_{\text{spanned}} = \left\{ (b_{(\ell)}, b'_{(\ell)}) \right\}_{\ell=1}^{k^{**}}, \tag{23}$$

and the set of anomalies classified as spanned is the union:

$$\mathcal{S}_{\text{joint}} = \bigcup_{\ell=1}^{k^{**}} \{ b_{(\ell)}, b'_{(\ell)} \}. \tag{24}$$

This procedure sharpens the inference by conditioning only on models where anomalies are either jointly unspanned or jointly spanned. It simplifies the model space, strengthens the evidence required for joint pricing, and controls the expected false discovery rate among selected pairs at level q.

2.3. Anomaly Spanning: Three at a Time

In our third step, we evaluate spanning for *triplets* of anomalies. This stage guards against higher–order masking (where no singleton or pair appears spanned, but the triplet is jointly spanned) and further de–duplicates family redundancy (multiple close proxies for the same risk). It also accommodates conditional spanning patterns that only emerge when all three signals are considered together. As in earlier steps, we impose a joint decision, either all three are in $\boldsymbol{w}$ (jointly spanned) or all three are in $\boldsymbol{x}$ (jointly unspanned) and control multiplicity by applying EFDR at the triplet level.

Let $c_1, \ldots, c_r$ denote the anomalies not classified as spanned after the two-at-a-time step, where $r = m - k^{**} = n - k^* - k^{**}$. For each triplet (c_i, c_k, c_ℓ) with $i < k < \ell$, we construct the augmented factor set:

$$\boldsymbol{f}^{(i,k,\ell)} = (\texttt{FF6}, c_i, c_k, c_\ell)$$

There are $J^{**} = 2^9 - 1 = 511$ nontrivial partitions of this 9-factor set into spanned ($\boldsymbol{w}_j^{(i,k,\ell)}$) and unspanned ($\boldsymbol{x}_j^{(i,k,\ell)}$) subsets. We restrict attention to the subset of models in which:

- **either** all three anomalies $\{c_i, c_k, c_\ell\}$ are in $\boldsymbol{w}_j^{(i,k,\ell)}$ (jointly spanned),
- **or** all three are in $\boldsymbol{x}_j^{(i,k,\ell)}$ (jointly unspanned).

We exclude models where only some of the three anomalies are spanned, to maintain

symmetry and eliminate ambiguous configurations. Each model $\mathbb{M}_j^{(i,k,\ell)}$ takes the form:

$$\boldsymbol{x}_j^{(i,k,\ell)} = \boldsymbol{\lambda}_j^{(i,k,\ell)} + \boldsymbol{u}_j^{(i,k,\ell)}, \tag{25}$$

$$\boldsymbol{w}_j^{(i,k,\ell)} = \Gamma_j^{(i,k,\ell)} \boldsymbol{x}_j^{(i,k,\ell)} + \boldsymbol{\varepsilon}_j^{(i,k,\ell)}, \tag{26}$$

where there is no intercept in the second equation.

Similarly, we still have $J_{\text{joint}} = 127$ admissible models, as shown in Table A.3. Among them, $\mathbb{M}_{120}^{(i,k,\ell)}$ and $\mathbb{M}_{127}^{(i,k,\ell)}$ denote, respectively, the cases where the tuple (c_i, c_k, c_ℓ) is jointly spanned by `FF6` and where it is not in the span of `FF6`.

Let $m_j^{(i,k,\ell)}$ denote the marginal likelihood of model $\mathbb{M}_j^{(i,k,\ell)}$, computed as in Chib (1995). Assuming a uniform prior over the restricted model space, we compute:

$$\Pr(\mathbb{M}_j^{(i,k,\ell)} \mid \text{data}) = \frac{m_j^{(i,k,\ell)}}{\sum_l m_l^{(i,k,\ell)}}. \tag{27}$$

We define the joint inclusion probability:

$$p_{i,k,\ell}^{\text{joint}} = \sum_{j:\{c_i,c_k,c_\ell\}\subset \boldsymbol{w}_j^{(i,k,\ell)}} \Pr(\mathbb{M}_j^{(i,k,\ell)} \mid \text{data}), \tag{28}$$

which is the posterior probability that the triplet is jointly spanned.

We test the hypotheses:

- Null hypothesis $H_0^{(i,k,\ell)}$: none of $\{c_i, c_k, c_\ell\}$ are spanned
- Alternative hypothesis $H_1^{(i,k,\ell)}$: all of $\{c_i, c_k, c_\ell\}$ are jointly spanned

The local false discovery rate for each triplet is:

$$\text{lfdr}_{i,k,\ell}^{\text{joint}} = 1 - p_{i,k,\ell}^{\text{joint}}. \tag{29}$$

Let $n^{**} = \binom{r}{3}$ denote the number of distinct anomaly triplets. We sort the local fdrs in

increasing order:

$$\text{lfdr}^{\text{joint}}_{(1)} \leq \text{lfdr}^{\text{joint}}_{(2)} \leq \cdots \leq \text{lfdr}^{\text{joint}}_{(n^{**})}. \tag{30}$$

Then compute for each $k = 1, \ldots, n^{**}$:

$$\text{EFDR}(k) = \frac{1}{k} \sum_{\ell=1}^{k} \text{lfdr}^{\text{joint}}_{(\ell)}. \tag{31}$$

We identify the largest k^{***} such that $\text{EFDR}(k^{***}) \leq q$. The top k^{***} triplets are declared as jointly spanned:

$$\mathcal{T}_{\text{spanned}} = \{(c_{(\ell)}, c'_{(\ell)}, c''_{(\ell)})\}_{\ell=1}^{k^{***}}, \tag{32}$$

and the set of anomalies classified as spanned is:

$$\mathcal{S}_{\text{triplet}} = \bigcup_{\ell=1}^{k^{***}} \{c_{(\ell)}, c'_{(\ell)}, c''_{(\ell)}\}. \tag{33}$$

This extension maintains symmetry, sharpens the inference, and ensures that only triplets with strong joint evidence of spanning are declared as priced, while controlling the expected false discovery rate at level q.

2.4. Anomaly Spanning: General Case

Now that we have established the set-up and provided details and intuition for how our method is implemented in particular cases, we describe the general case when `FF6` is paired with r anomalies. Let $c_1, \ldots, c_r$ denote the anomalies remaining after the previous step. For each subset $\mathcal{S} \subset \{c_1, \ldots, c_r\}$ such that $|\mathcal{S}| = s$, we construct the augmented factor set:

$$\boldsymbol{f}^{(\mathcal{S})} = (\texttt{FF6}, c_{i_1}, \ldots, c_{i_s}) \quad \text{for } \mathcal{S} = \{c_{i_1}, \ldots, c_{i_s}\}. \tag{34}$$

There are $J_s = 2^{6+s} - 1$ nontrivial partitions of this factor set into spanned and unspanned subsets. We restrict attention to models where the anomalies in $\mathcal{S}$ are either:

- all in the spanned set $\boldsymbol{w}_j^{(\mathcal{S})}$ (jointly spanned), or

- all in the unspanned set $\boldsymbol{x}_j^{(\mathcal{S})}$ (jointly unspanned).

This restriction eliminates mixed configurations and simplifies the classification. Each model $\mathbb{M}_j^{(\mathcal{S})}$ in the restricted class takes the form:

$$\boldsymbol{x}_j^{(\mathcal{S})} = \boldsymbol{\lambda}_j^{(\mathcal{S})} + \boldsymbol{u}_j^{(\mathcal{S})}, \tag{35}$$

$$\boldsymbol{w}_j^{(\mathcal{S})} = \Gamma_j^{(\mathcal{S})} \boldsymbol{x}_j^{(\mathcal{S})} + \boldsymbol{\varepsilon}_j^{(\mathcal{S})}, \tag{36}$$

where no intercept appears in the second equation.

Let $m_j^{(\mathcal{S})}$ denote the marginal likelihood of model $\mathbb{M}_j^{(\mathcal{S})}$, computed using the method of Chib (1995). With a uniform prior over the restricted model space, the posterior model probabilities are:

$$\Pr(\mathbb{M}_j^{(\mathcal{S})} \mid \text{data}) = \frac{m_j^{(\mathcal{S})}}{\sum_k m_k^{(\mathcal{S})}}. \tag{37}$$

Define the joint inclusion probability for the subset $\mathcal{S}$ as:

$$p_{\mathcal{S}}^{\text{joint}} = \sum_{j:\mathcal{S}\subset \boldsymbol{w}_j^{(\mathcal{S})}} \Pr(\mathbb{M}_j^{(\mathcal{S})} \mid \text{data}). \tag{38}$$

This is the posterior probability that all anomalies in $\mathcal{S}$ are jointly spanned.

We now test the hypotheses:

- Null hypothesis $H_0^{(\mathcal{S})}$: none of the anomalies in $\mathcal{S}$ are spanned
- Alternative hypothesis $H_1^{(\mathcal{S})}$: all anomalies in $\mathcal{S}$ are jointly spanned

The local false discovery rate for each subset is:

$$\text{lfdr}_{\mathcal{S}} = 1 - p_{\mathcal{S}}^{\text{joint}}. \tag{39}$$

Let $n_s = \binom{r}{s}$ denote the number of distinct s-anomaly subsets. We sort the lfdrs:

$$\text{lfdr}_{(1)} \leq \text{lfdr}_{(2)} \leq \cdots \leq \text{lfdr}_{(n_s)}. \tag{40}$$

For each $k = 1, \ldots, n_s$, compute:

$$\text{EFDR}(k) = \frac{1}{k} \sum_{\ell=1}^{k} \text{lfdr}_{(\ell)}. \tag{41}$$

Let k^* be the largest index such that $\text{EFDR}(k^*) \leq q$, for a pre-specified threshold $q \in (0, 1)$. The top k^* subsets are declared as jointly spanned:

$$\mathcal{G}_{\text{spanned}} = \{\mathcal{S}_{(\ell)}\}_{\ell=1}^{k^*}, \tag{42}$$

and the anomalies classified as spanned are:

$$\mathcal{S}_{\text{final}} = \bigcup_{\ell=1}^{k^*} \mathcal{S}_{(\ell)}. \tag{43}$$

This procedure generalizes the two- and three-anomaly steps, maintains symmetry, and controls the expected false discovery rate at level q. We continue this iterative process until the set of unspanned anomalies stabilizes.

Building on Proposition 1, which considers EFDR control at the individual anomaly level, we now establish a second decision-theoretic optimality result about our Bayesian screening method. We show that our sequential screening across nested anomaly sets maintains global EFDR control.

Proposition 2 (Sequential EFDR control across stages)**.** *Fix data $\mathcal{D}$. For each candidate subset $\mathcal{S}$ of anomalies at stage s, let $p_{\mathcal{S}}^{joint} = \Pr(H_{1,\mathcal{S}} \mid \mathcal{D})$ denote the posterior probability that the set $\mathcal{S}$ is* jointly spanned *(alternative true), and define the local false discovery rate $\text{lfdr}_{\mathcal{S}} = 1 - p_{\mathcal{S}}^{joint}$. Let $\mathcal{H}_s$ be the collection of subsets tested at stage s and write $\text{lfdr}_{(1)}^{(s)} \leq \cdots \leq \text{lfdr}_{(n_s)}^{(s)}$ for the ordered lfdr values over $\mathcal{H}_s$.*

At level $q \in (0, 1)$, the Bayes–optimal EFDR step–up rule (Prop. 1) selects

$$k_s^\star = \max\left\{ k \in \{0, \ldots, n_s\} : \frac{1}{k} \sum_{\ell=1}^{k} \text{lfdr}_{(\ell)}^{(s)} \leq q \right\}, \qquad \mathcal{S}_s^\star = \left\{ \mathcal{S}_{(1)}^{(s)}, \ldots, \mathcal{S}_{(k_s^\star)}^{(s)} \right\},$$

and we define the associated posterior–null threshold

$$\tau_s \;\equiv\; \mathrm{lfdr}^{(s)}_{(k_s^\star)} \quad \textit{(with the convention that } \tau_s = +\infty \textit{ if } k_s^\star = 0)\,.$$

so that the same selection can be written equivalently as

$$\mathcal{S}_s^\star \;=\; \big\{\, \mathcal{S} \in \mathcal{H}_s \;:\; \mathrm{lfdr}_{\mathcal{S}} \le \tau_s \,\big\}.$$

Assume nested elimination across stages:

$$\mathcal{H}_{s+1} \subseteq \mathcal{H}_s \setminus \mathcal{S}_s^\star,$$

so that the selected sets $\mathcal{S}_1^\star, \ldots, \mathcal{S}_S^\star$ are pairwise disjoint. Let the cumulative discovery set be

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

Then the global EFDR computed at the subset level *satisfies*

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

Proof. All probabilities are posterior given $\mathcal{D}$. For any finite selection $\mathcal{A}$ of subsets, the EFDR at the subset level is

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

By Proposition 1, the stage–s step–up rule yields $k_s^\star$ and $\mathcal{S}_s^\star$ such that

$$\frac{1}{\max\big(|\mathcal{S}_s^\star|, 1\big)} \sum_{\mathcal{S} \in \mathcal{S}_s^\star} \mathrm{lfdr}_{\mathcal{S}} \;\le\; q\,.$$

Because $\mathcal{H}_{s+1} \subseteq \mathcal{H}_s \setminus \mathcal{S}_s^\star$, the selections $\mathcal{S}_1^\star, \ldots, \mathcal{S}_S^\star$ are disjoint, so

$$\sum_{\mathcal{S}\in\mathcal{S}^\star} \mathrm{lfdr}_{\mathcal{S}} \;=\; \sum_{s=1}^{S} \sum_{\mathcal{S}\in\mathcal{S}_s^\star} \mathrm{lfdr}_{\mathcal{S}} \;\le\; \sum_{s=1}^{S} q\cdot \max\big(|\mathcal{S}_s^\star|, 1\big)\,.$$

If at least one stage is nonempty, then $|\mathcal{S}^\star| = \sum_{s=1}^{S} |\mathcal{S}_s^\star| \ge 1$, and hence

$$\mathrm{EFDR}\big(\mathcal{S}^\star \mid \mathcal{D}\big) \;=\; \frac{\sum_{s=1}^{S}\sum_{\mathcal{S}\in\mathcal{S}_s^\star} \mathrm{lfdr}_{\mathcal{S}}}{\sum_{s=1}^{S} |\mathcal{S}_s^\star|} \;\le\; \frac{\sum_{s=1}^{S} q\,|\mathcal{S}_s^\star|}{\sum_{s=1}^{S} |\mathcal{S}_s^\star|} \;=\; q\,.$$

If every $\mathcal{S}_s^\star = \varnothing$, then $\mathcal{S}^\star = \varnothing$ and $\mathrm{EFDR}(\mathcal{S}^\star \mid \mathcal{D}) = 0 \le q$ by convention. This proves the claim. □

Proposition 2 formalizes that the sequential screening and elimination process preserves EFDR control globally, not just within each stage. Intuitively, previously discovered (spanned) sets are removed from subsequent testing, ensuring that the posterior expected proportion of false discoveries in the cumulative set of spanned anomalies remains bounded by q. This property provides strong decision-theoretic guarantees for our multi-stage Bayesian screening method.

3. Empirical Study

We obtain U.S. monthly data for the key risk factors, `Mkt`, `SMB`, `HML`, `RMW`, `CMA`, and `MOM`, from the Fama-French data library. Our analysis builds on the database constructed by Jensen, Kelly, and Pedersen (2023), which contains 153 factors grouped into 13 thematic categories. We use monthly value-weighted returns from this dataset and restrict the sample period to January 1985 through December 2024 to ensure full coverage across all 153 factors. This results in a total sample size of $T = 480$ months. The training sample comprises the first 25% of observations ($T_{\text{tr}} = 120$), corresponding to the first 10 years, with the remaining $T_{\text{est}} = 360$ observations from Jan 1995 to Dec 2024 used for model estimation [3]. We refer to the $\boldsymbol{x}$ and $\boldsymbol{w}$ at each step as *remaining anomalies* and *fake anomalies*, respectively.

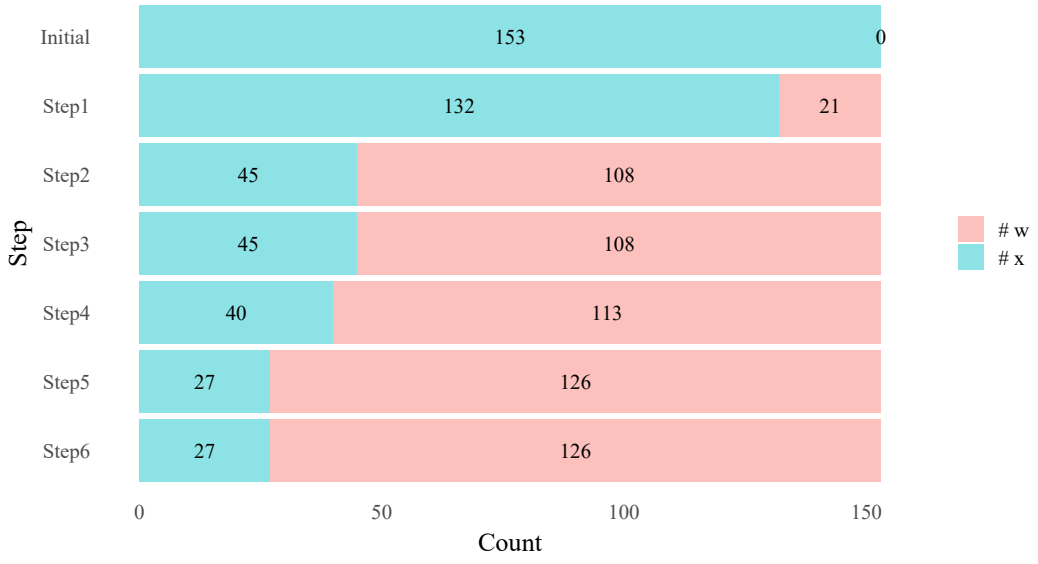
[3] In the non-Bayesian empirical exercises, we continue to use a full sample of $T = 480$ months.

3.1. Results

We begin with a straightforward question: when FF6 is taken as the benchmark, which are the *fake anomalies*, that is, anomalies that can be spanned jointly by FF6 and a small number of additional anomalies. To address this, we implement our stepwise methodology based on EFDR ($q = 0.2$) across the initial set of 153 candidate factors. At Step s, FF6 is allowed to enter jointly with s candidate anomalies. Once an anomaly or a group of anomalies is classified as spanning in the preceding step, these are recorded as fake and removed from the candidate pool, while the remaining factors proceed to the next step.

Figure 1 reports the identification path. Steps 1, 2, 4, and 5 contribute 21, 87, 5, and 13 new fake anomalies, respectively, while Step 3 yields none. By the end of Step 5, the classification stabilizes at 126 fake anomalies and 27 remaining factors, accounting for 82.4% and 17.6% of the total candidate anomalies set, respectively. This pattern indicates that, within the class of "FF6 plus a small number of anomalies" models, most candidate factors are explained, while only a small fraction provides genuinely incremental pricing information.

Figure 1 Counts of Fake Anomalies and Remaining Anomalies.
Note: This figure reports the counts of anomalies classified as $\boldsymbol{w}$ and $\boldsymbol{x}$ by our procedure, using FF6 as the benchmark. Green bars denote remaining anomalies ($\boldsymbol{x}$), and red bars denote fake anomalies ($\boldsymbol{w}$).



From an economic perspective, the identification path offers clear intuition. FF6 already captures a substantial portion of systematic risk exposures in the cross section. Once a few additional anomaly factors are allowed, many candidates exhibit informational redundancy or substitutability with these benchmark exposures and are classified as $\boldsymbol{w}$. The 27 remaining anomalies are more likely to reflect exposures heterogeneous to FF6, thereby carrying genuinely incremental pricing power.

It is essential to note that $q = 0.2$ is not a strict threshold, but rather a relatively lenient criterion for the candidate factors. Figure 2 displays the barplot of lfdr and EFDR in Step 1. Even though the maximum EFDR in this step is about 0.5, choosing $q = 0.2$ corresponds to a relatively low quantile, meaning that the factors classified as $\boldsymbol{w}$ in step 1 are screened under a fairly loose standard. At the same time, setting $q = 0.2$ also preserves computational tractability; otherwise, in later steps the number of models to be evaluated can easily run into the tens of millions. Figure A.1 illustrates the number of $\boldsymbol{w}$ factors that can be identified from the EFDR procedure at $q = 0.2$ when alternative thresholds are applied. We find that $q = 0.2$ yields substantially more $\boldsymbol{w}$ in the first step compared with $q = 0.1$ and $q = 0.15$, and it continues to identify additional $\boldsymbol{w}$ in Steps 4 and 5, thereby improving the overall screening efficiency. For robustness, we also consider the procedure with $q = 0.1$ using FF6 as the benchmark. As shown in Figure A.2, when q ranges from 0.10 to 0.20, the final number of identified $\boldsymbol{w}$ is stable, with only minor differences in the composition of the sets.

To further illustrate the stepwise selection process, Figure 3 displays the EFDR and lfdr distributions for each step. Statistically, the absence of new $\boldsymbol{w}$ in Step 3 reflects a selection-exhaustion effect: Steps 1-2 have already removed the anomalies most easily spanned. In Steps 1 and 2, EFDR values fall primarily between 0.1 and 0.6, so adopting $q = 0.2$ as the threshold identifies a subset of $\boldsymbol{w}$ factors. From Step 3 onward (panels (c)–(f)), however, both EFDR and lfdr concentrate near 1 across candidate combinations. As the set of explained $\boldsymbol{w}$ expands and the candidate pool contracts, locating additional $\boldsymbol{w}$ among the remaining anomalies becomes increasingly difficult, accounting for the sharp drop in incremental detections at later steps.

Figure 2 Bar plot of EFDR and lfdr in the first step.

Note: This figure plots EFDR and lfdr for each anomaly in Step 1 of our procedure. The red bars show EFDR, and the green bars show lfdr.

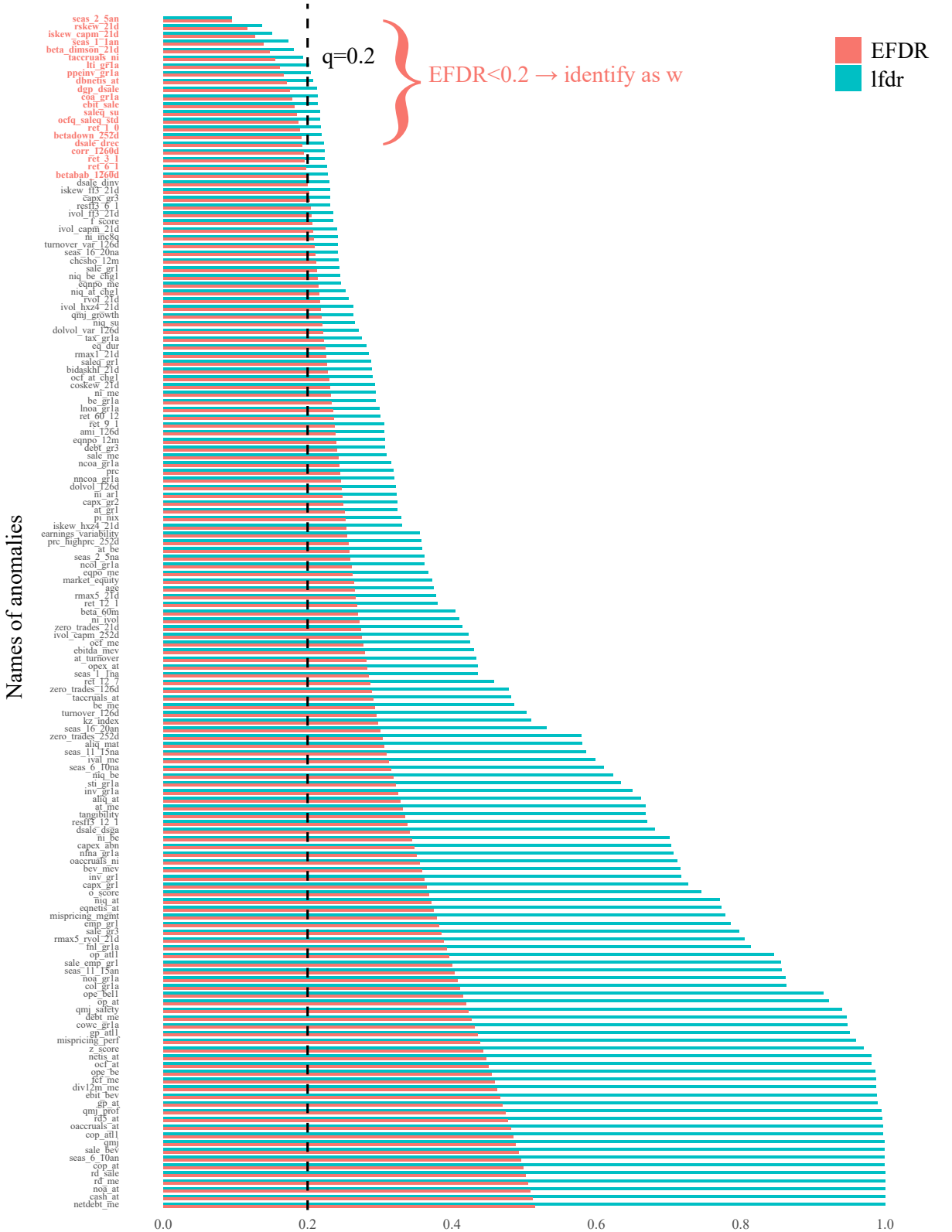
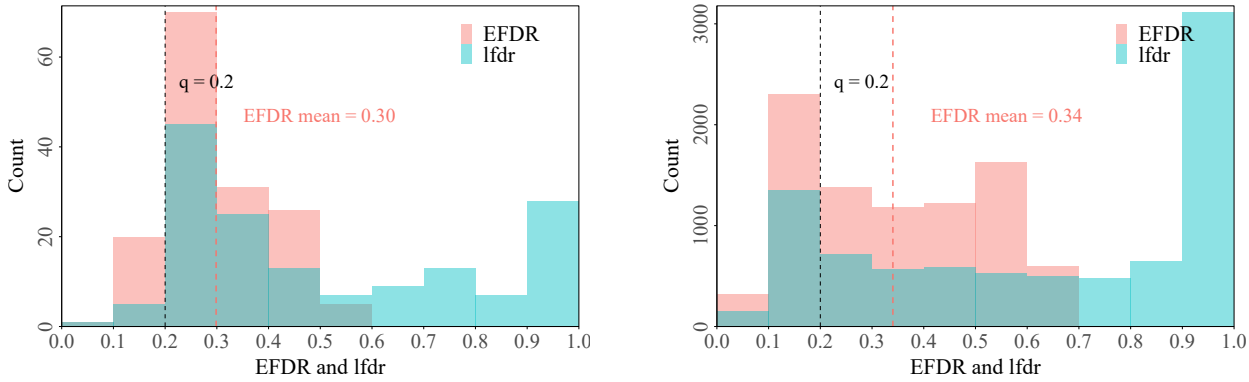
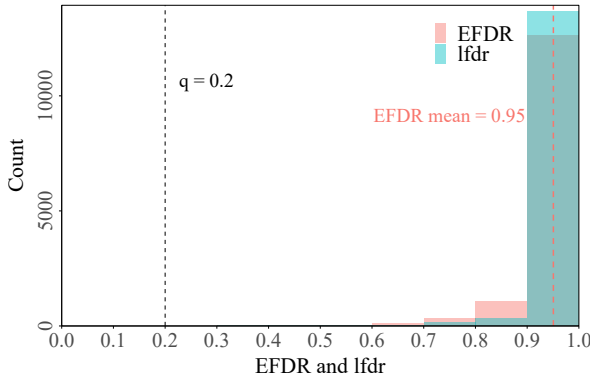


Figure 3 Histogram of EFDRs and lfdrs across Steps 1–6.
Note: This figure displays, for steps 1–6 under the `FF6` benchmark, overlaid histograms of EFDR (red) and lfdr (green). We also mark the threshold at $q = 0.2$ and report the mean EFDR in each panel.

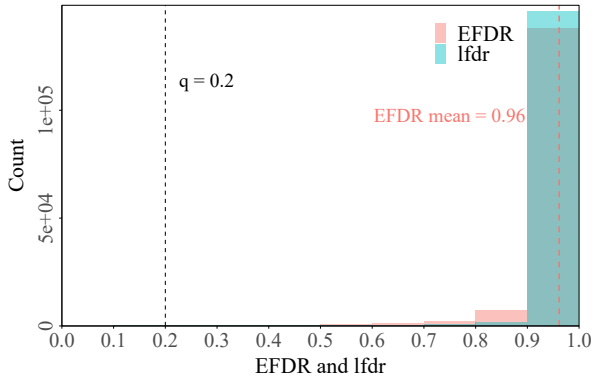


(a) Step 1

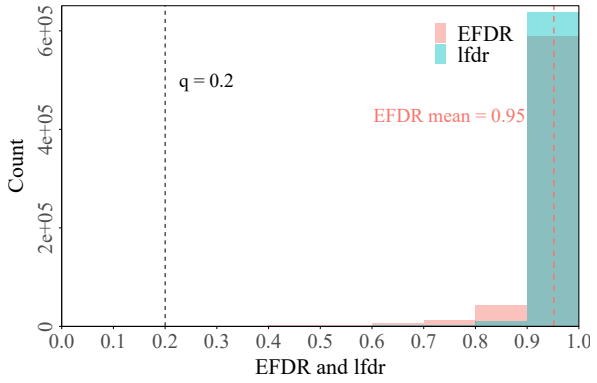
(b) Step 2



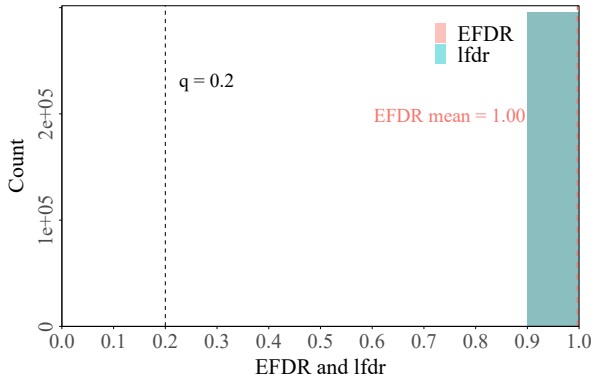
(c) Step 3



(d) Step 4



(e) Step 5



(f) Step 6

3.2. Why these?

Using FF6 as the benchmark effectively lays down the core pricing exposures ex ante. Table 1 shows that once these six dimensions are in the span, themes that are economically homogeneous with them collapse into the spanned set $\boldsymbol{w}$. The concentration is stark for Investment, Momentum, and Low risk (100% in $\boldsymbol{w}$ for each). Investment signals (asset growth, capex, inventory buildup, and related net operating assets (NOA)-based growth measures) are largely spanned by the CMA; momentum variants load on MOM; and low-risk constructions are well approximated by combinations of FF6 factors that reproduce lower market beta, larger value/quality tilts, and hence are readily spanned.

Themes adjacent to other FF6 legs also largely migrate into $\boldsymbol{w}$. Profit growth (91.7% in $\boldsymbol{w}$) is closely tied to the earnings-based profitability leg RMW; Debt issuance (85.7%) comoves with investment and value through dilution and external-financing cycles, allowing CMA and HML to absorb much of the variation. Value (83.3%) and Size (80.0%) are primarily absorbed by HML and SMB, respectively. Seasonality (83.3%) captures predictable short-horizon return patterns driven by calendar effects and institutional rebalancing, which makes it closely related to momentum (MOM). Short-term reversal (80%) captures microstructure- and liquidity-driven short-horizon price reversals, which makes it negatively aligned with MOM and often accompanied by mild exposure to size (SMB). In such case, the economic mechanism behind the characteristic is already encoded in at least one FF6 factor, so a no-intercept spanning relation is easy to establish.

By contrast, themes that only partially overlap with FF6 leave more survivors in $\boldsymbol{x}$. Quality (58.8% in $\boldsymbol{w}$) and Profitability (63.6%) capture cash-flow durability and operating efficiency, which RMW, built on operating profitability rather than cash, does not fully span. Accruals (50.0%) and related working-capital/NOA constructs primarily capture accounting recognition/timing effects and the composition of operating investment. These are only partially related to the "quantity of investment" in CMA. Low leverage (63.6%) primarily reflects financing frictions in the price and quantity of external capital, while also embedding profitability/safety, cash buffers, collateral/intangibles, and trading liquidity.

Table 1 Number and Percentage of Anomalies.

Note: This table reports, by theme, the number and percentage of fake anomalies. Themes are ordered in ascending order by the number of remaining 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%

Consequently, it is only imperfectly proxied by `HML`, `SMB`, or `CMA`. The result is a lower w-share and a nontrivial x-residual along these dimensions.

Overall, a natural regularity emerges: the closer a factor's economic content is to the benchmark, the more likely it is to be spanned and thus classified as w. By contrast, the x factors that deliver genuinely incremental pricing power tend to concentrate in economic channels that are more heterogeneous relative to the benchmark.

4. Evaluating Fake Anomalies and Last-stage Anomalies

In Section 3.1, we use `FF6` as the benchmark and identify 126 w and 27 x. To distinguish them, we refer to the 126 w as *fake anomalies* and the 27 x as *last-stage anomalies*. In fact, the process of applying our method in Section 3.1 is akin to "*gold panning*": the sand (fake anomalies) is filtered out, and the remaining portion contains both gold and sand. In other words, the last-stage anomalies are either genuine factors or not, a fact we

cannot ascertain since the true SDF is unknown. What we can assert, however, is that up to the level of our EFDR control, 126 anomalies can be spanned by FF6 and combinations of the remaining anomalies, and that, therefore, these can be classified as fake anomalies.

Furthermore, in this section, we demonstrate the differences between $\boldsymbol{x}$ and $\boldsymbol{w}$ through a series of empirical exercises. First, following Chib et al. (2024), we evaluate the out-of-sample (OOS) predictive performance of different factor combinations by utilizing the predictive likelihood, which is the standard Bayesian predictive measure[4]. Next, we compare the out-of-sample Sharpe ratios and pricing performance across different combinations of $\boldsymbol{x}$ and $\boldsymbol{w}$. We will use three test-asset universes: the Fama-French 49 industry portfolios, 100 P-Tree portfolios (Cong et al., 2025), and 120 ($2 \times 3 \times 20$) bi-sort portfolios constructed from 20 characteristics[5].

As discussed in Section 3.1, the number of selected $\boldsymbol{x}$ stabilizes after the fifth step of the joint selection procedure. Accordingly, we first select five $\boldsymbol{x}$ factors from the set of 27, yielding $\binom{27}{5}$ =80,730 possible combinations. For each combination, we compute the log average predictive likelihood (LAPL) for $\boldsymbol{f} = 5\boldsymbol{x}$ based on Markov chain Monte Carlo (MCMC) samples, thereby obtaining a distribution of LAPL values. Analogously, we randomly draw 50,000 combinations of $5\boldsymbol{w}$ from the 127 fake anomalies and perform the same calculation. We then augment these combinations with FF6. The computation is implemented following Chib et al. (2024). [6]

Table 2 reports summary statistics of the LAPL across these different factor-combination classes. Across both the anomalies-only specifications and the anomalies-plus-FF6 specifications, factors identified as fake anomalies deliver lower predictive

[4]For a given realization of future data, the predictive likelihood, like the marginal likelihood, reduces to a scalar and thus provides a natural basis for ranking models across the candidate model space.

[5]The details of the characteristics are shown in Table A.4.

[6]Specifically, we implement this procedure by partitioning the 40-year sample into 30 years for estimation and 10 years for OOS prediction; the first 10 years are used to fix the prior, as discussed earlier. We then consider 9 contiguous in-sample/OOS splits of the available data. For each split, we re-estimate the model and compute the predictive likelihood. The performance metric is the average of these 9 predictive likelihoods. We repeat this computation for every factor combination. For each factor combination, we consider only the model space that corresponds to that combination. The subsequent Bayesian empirical analyses follow the same setup.

likelihood on average. For example, in Panel B the mean LAPL for $(5\boldsymbol{x}, \texttt{FF6})$ exceeds that for $(5\boldsymbol{w}, \texttt{FF6})$ by about 24, indicating that models built with $(5\boldsymbol{x}, \texttt{FF6})$ outperform the competing models on the predictive dimension.

Table 2 Statistics of Log Average of Predictive Likelihoods for Different Factor Combinations.

Note: This table reports the statistics of log average of predictive likelihood for selected factor combinations. We consider model classes formed by combining $\boldsymbol{x}$ or $\boldsymbol{w}$ with/without the benchmark. Specifically, $(5\boldsymbol{x}, \texttt{FF6})$ denotes all 80,730 models obtained by augmenting FF6 with 5 $\boldsymbol{x}$ in turn; for each model, we compute the corresponding LAPL and then construct the reported statistic. Since the number of $\boldsymbol{w}$ combinations can be very large, we randomly sample 50,000 combinations from the feasible set.

$\boldsymbol{f}$	# Factor	# Comb	Mean	Std.	Median	2.5% qtile	97.5% qtile
Panel A. Only Anom.							
$5\boldsymbol{x}$	5	80,730	1160.33	54.44	1154.43	1070.12	1192.69
$5\boldsymbol{w}$	5	50,000	1117.52	61.30	1115.46	1003.55	1157.42
Panel B: Anom. & Bench. (Match with Method)							
$(5\boldsymbol{x}, \texttt{FF6})$	11	80,730	2500.04	57.63	2494.47	2403.08	2534.96
$(5\boldsymbol{w}, \texttt{FF6})$	11	50,000	2476.34	56.28	2475.87	2368.04	2513.06

Having established that a set of $\boldsymbol{w}$ is less effective than a set of $\boldsymbol{x}$ from a predictive perspective, we next examine the performance of different factor combinations. Specifically, we calculate the annualized Sharpe ratios (SR) of tangency portfolios constructed from the identified $\boldsymbol{x}$ and $\boldsymbol{w}$, both full-sample and out-of-sample (using half of the whole data). The OOS SR is obtained by applying the portfolio weights estimated in-sample to out-of-sample realized returns. Analogous to Table 2, we enumerate all $\binom{27}{5}$ possible combinations and report the average. For comparison, we randomly select five fake anomalies from the set of 126 and compute the corresponding full-sample and out-of-sample SRs. In addition, we examine the performance of the ($\boldsymbol{x}$, FF6) specification and the corresponding number of $\boldsymbol{w}$ factors.

Table 3 shows that the $\boldsymbol{x}$ remaining anomalies identified by our method consistently deliver higher Sharpe ratios both full-sample and out-of-sample. Moreover, the combination of "five last-stage anomalies and FF6" achieves the best out-of-sample performance. This finding follows from the design of our procedure. The combinations

align with its structure: when FF6 serves as the benchmark, the resulting $\boldsymbol{x}$ are exactly those selected in the five-at-a-time step.

Table 3 Average Sharpe Ratios of Tangency Portfolios Across Different Factors.

Note: This table reports the average tangency-portfolio Sharpe ratio across the full sample and out-of-sample window, for various factor combinations. We evaluate three configurations for remaining anomalies: (i) choose 5 $\boldsymbol{x}$ from the 27 $\boldsymbol{x}$; (ii) choose 5 $\boldsymbol{x}$ from the 27 $\boldsymbol{x}$ and augment the benchmark (FF6); and (iii) use all $\boldsymbol{x}$ in combination with the benchmark (FF6). For $\boldsymbol{w}$, the comparison uses sets whose cardinality matches the number of $\boldsymbol{x}$ in each configuration; specifically, we randomly draw 50,000 combinations from the 126 $\boldsymbol{w}$. We use the second half of the full sample for OOS evaluation.

$\boldsymbol{f}$	# Factor	# Comb	Full sample	OOS
Panel A. Only Anom.				
$5\boldsymbol{x}$	5	80,730	0.98	0.86
$5\boldsymbol{w}$	5	50,000	0.60	0.24
Panel B: Anom. & Bench. (Match with Method)				
$(5\boldsymbol{x}, \texttt{FF6})$	5+6	80,730	1.71	**1.18**
$(5\boldsymbol{w}, \texttt{FF6})$	5+6	50,000	1.38	0.69
Panel C: Anom. and Bench.				
$(27\boldsymbol{x}, \texttt{FF6})$	27+6	1	**2.26**	1.01
$(27\boldsymbol{w}, \texttt{FF6})$	27+6	50,000	1.93	0.48

Furthermore, we evaluate the ability of the 27 last-stage anomalies to explain the cross section of expected returns across various sets of test assets. Given that Table 3 shows the combination of "five $\boldsymbol{x}$ plus FF6" performs particularly well, and this specification aligns with our methodology, we focus on reporting the results for all such combinations along with the corresponding number of $\boldsymbol{w}$ factors.

Table 4 reports the average explanatory power across different factors and test assets in the full sample and out-of-sample. Performance is summarized using several metrics: Root Mean Square Alpha (RMS α), Root Mean Squared Error (RMSE), absolute value of alpha ($|\alpha|$), the GRS statistic, and Total R^2(%) [7]. All reported metrics are computed as the averages of the corresponding metrics across all possible combinations. Specifically,

[7] RMS $\alpha = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_i^2}$, RMSE $= \sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}\hat{\varepsilon}_{it^2}}$, $|\alpha| = \frac{1}{N}\sum_{i=1}^{N}|\hat{\alpha}_i|$, $\text{TR}^2 = 1 - \frac{\sum_{i=1}^{N}\sum_{i=1}^{T}\hat{\varepsilon}_{it^2}}{\sum_{i=1}^{N}\sum_{i=1}^{T}r_{it^2}}$.

Table 4 Average Explanatory Power of Different Factors Across Test Assets.

Note: This table reports the average asset-pricing performance across the out-of-sample window, for (i) all factor sets formed by augmenting FF6 with every possible combination of 5 last-stage anomalies and (ii) 50,000 sets formed by augmenting FF6 with random combinations of 5 fake anomalies. We use the second half of the full sample for OOS evaluation.

$\boldsymbol{f}$	# Factor	# Comb	Metrics				
			RMS α	RMSE	$\lvert\alpha\rvert$	GRS stat.	Total R^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

across test-asset universes, factor sets built from the remaining anomalies in conjunction with FF6 outperform those built from the fake anomalies. In particular, the ($5\boldsymbol{x}$, FF6) combinations exhibit smaller mispricing in both the full sample and the out-of-sample period. For example, considering the full sample, in Panel A, the RMS alpha equals 0.0041 versus 0.0043 for $\boldsymbol{w}$, and a stronger overall fit: in all three panels, $(5\boldsymbol{x}, \texttt{FF6})$ attains higher TR^2 than the corresponding $\boldsymbol{w}$ sets. The evidence indicates that the $\boldsymbol{x}$ factors identified relative to the FF6 benchmark exhibit stronger pricing ability than the $\boldsymbol{w}$ fake anomalies.

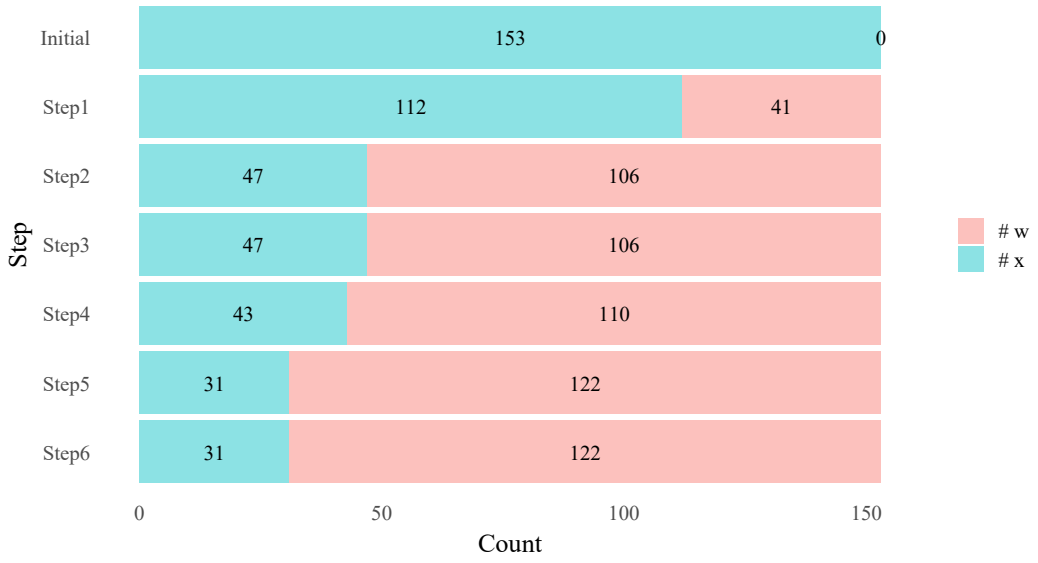
4.1. Fake Anomalies under Different Benchmarks

In Section 4, we have shown that the $\boldsymbol{w}$ identified by our procedure under FF6 benchmark can be priced by the benchmark in conjunction with the remaining anomalies, and that $\boldsymbol{w}$ and $\boldsymbol{x}$ exhibit pronounced differences in pricing strength. Naturally, the choice of benchmark may influence which $\boldsymbol{w}$ and $\boldsymbol{x}$ are ultimately identified. This arises because different benchmarks contain spanning information. However, our method is relatively benchmark-robust since the spanning set does not rely on the optimality of the starting

benchmark set.

To examine this, we reapply our procedure with another benchmark. Specifically, we retain the excess market return (Mkt) but replace the other FF6 factors with the profitability (ROE) and investment (IA) factors from the Hou, Xue, and Zhang (2020) q-factor model, the Pástor and Stambaugh (2003) liquidity factor (LIQ), the Frazzini and Pedersen (2014) betting-against-beta factor (BAB), and an alternative value factor, HMLD, proposed by Asness and Frazzini (2013). In other words, we continue to use six factors as the benchmark; apart from Mkt, all constituents differ from FF6. We denote this new benchmark by NewB6. Hence, the overall benchmark set is largely aligned with Chib, Zeng, and Zhao (2020). [8]

Figure 4 Counts of Fake Anomalies and Remaining Anomalies (Benchmark: NewB6).
Note: This figure reports the counts of anomalies classified as $\boldsymbol{w}$ and $\boldsymbol{x}$ by our procedure, using NewB6 as the benchmark. Green bars denote remaining anomalies ($\boldsymbol{x}$), and red bars denote fake anomalies ($\boldsymbol{w}$).



The results shown in Figure 4 indicate that, compared with Figure 1, when NewB6 is used as the benchmark, the evolution of screening efficiency is similar to that under FF6. One difference is that NewB6 identifies a larger number of spanned anomalies.

To further assess how benchmark choice affects the inference of spanned anomalies,

[8] Chib et al. (2020) also includes the QMJ factor, but since QMJ is among the 153 candidate anomalies in our sample, we do not include it in the benchmark here.

Figure 5 Percentage of Fake Anomalies within each Theme.
Note: This figure reports, for each theme category, the percentage of fake anomalies under two benchmarks: `FF6` (green) and `NewB6` (red). The y-axis lists category names together with the number of candidate anomalies in each category.

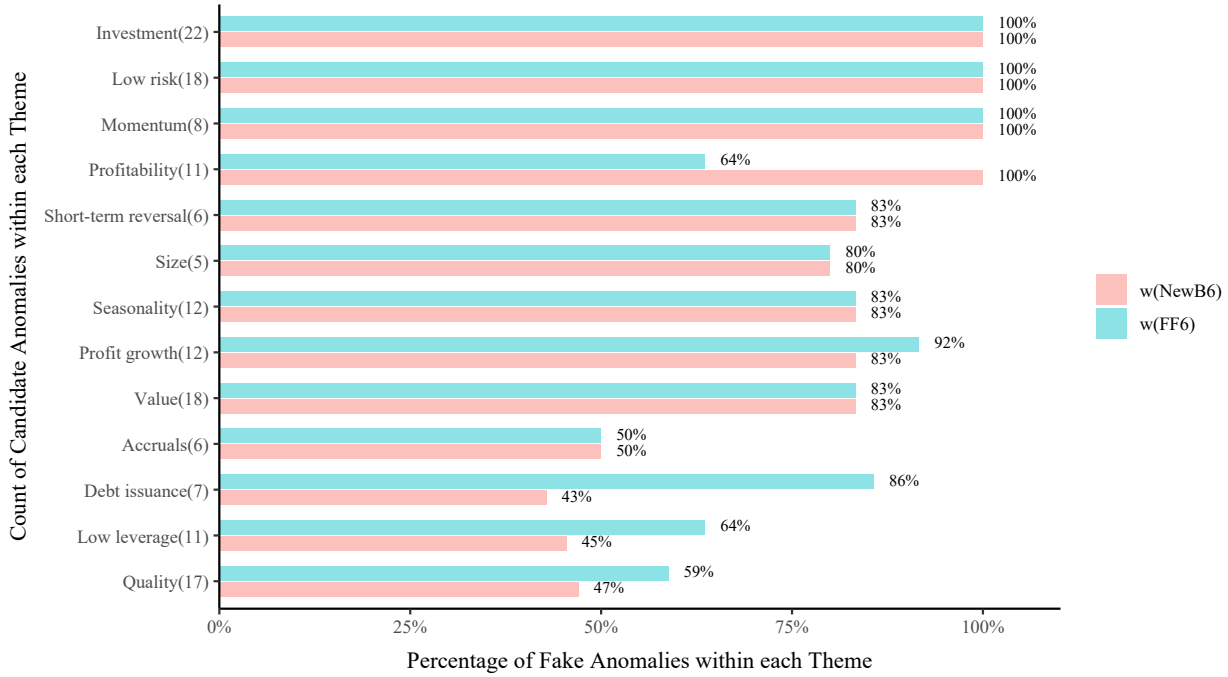


Figure 5 reports the percentage of fake anomalies within each economic category. For Investment, Low risk, and Momentum, the anomalies are classified as spanned regardless of the benchmark. The largest shifts occur for Profitability and Debt issuance. With `NewB6` as the benchmark, all candidates in Profitability are classified as spanned, largely because `ROE` in `NewB6` already embeds profitability information. By contrast, within Debt issuance, the share classified as spanned falls from 86% to 43% (from six to three). This decline reflects that the `NewB6` factors are not perfect substitutes for `CMA`, `SMB` and `MOM`, so they do not fully capture excess returns tied to investment quality (e.g., `capex_abn`) and financing risk (e.g., `fnl_gr1a`).

In fact, although the benchmark matters, the spanned anomalies identified under different benchmarks are highly similar. To further illustrate this phenomenon, Table 5 reports the degree of overlap in $\boldsymbol{w}_{\texttt{FF6}}$ and $\boldsymbol{w}_{\texttt{NewB6}}$. We consider four metrics, which summarize overlap from distinct angles: set similarity (Jaccard index), the precision

Table 5 Degree of Overlap in Fake Anomalies Under Different Benchmarks.

Note: This table measures the overlap between fake anomalies obtained under the FF6 and NewB6 benchmarks using several complementary metrics. # $\boldsymbol{a}$ denote the number of candidate anomalies in each theme; # $\boldsymbol{w}_{\texttt{FF6}}$ and # $\boldsymbol{w}_{\texttt{NewB6}}$ denote the numbers of fake anomalies under FF6 and NewB6, respectively; and # $\boldsymbol{w}_{\text{Inter}}$ denote the number identified as fake by both benchmarks. Jaccard index, $\text{Precision}_{\texttt{FF6}}$, $\text{Precision}_{\texttt{NB6}}$ and F_1-score are four metrics to describe the level of overlap.

Theme	# $\boldsymbol{a}$	# $\boldsymbol{w}_{\texttt{FF6}}$	# $\boldsymbol{w}_{\texttt{NewB6}}$	# $\boldsymbol{w}_{\text{Inter}}$	Jaccard index	$\text{Precision}_{\texttt{FF6}}$	$\text{Precision}_{\texttt{NB6}}$	F_1-score
Investment	22	22	22	22	1.00	1.00	1.00	1.00
Low risk	18	18	18	18	1.00	1.00	1.00	1.00
Seasonality	12	10	10	10	1.00	1.00	1.00	1.00
Momentum	8	8	8	8	1.00	1.00	1.00	1.00
Short-term reversal	6	5	5	5	1.00	1.00	1.00	1.00
Size	5	4	4	4	1.00	1.00	1.00	1.00
Accruals	6	3	3	3	1.00	1.00	1.00	1.00
Profit growth	12	11	10	10	0.91	0.91	1.00	0.95
Value	18	15	15	14	0.88	0.93	0.93	0.93
Low leverage	11	7	5	5	0.71	0.71	1.00	0.83
Profitability	11	7	11	7	0.64	1.00	0.64	0.78
Quality	17	10	8	7	0.64	0.70	0.88	0.78
Debt issuance	7	6	3	3	0.50	0.50	1.00	0.67
Total	153	126	122	116	0.88	0.92	0.95	0.94

of FF6 against NewB6 ($\text{Precision}_{\texttt{FF6}}$), the precision of NewB6 against FF6 ($\text{Precision}_{\texttt{NewB6}}$), [9] and their harmonic mean (F_1-score). Let # $\boldsymbol{w}_{\texttt{FF6}}$ and # $\boldsymbol{w}_{\texttt{NewB6}}$ denote the numbers of fake anomalies under FF6 and NewB6, respectively; and # $\boldsymbol{w}_{\text{Inter}}$ denote the number identified as fake by both benchmarks, i.e., $\boldsymbol{w}_{\text{Inter}} = \boldsymbol{w}_{\texttt{FF6}} \cap \boldsymbol{w}_{\texttt{NewB6}}$. The metrics are computed as follows:

$$\text{Jaccard } J = \frac{\#\boldsymbol{w}_{\text{Inter}}}{\#\boldsymbol{w}_{\texttt{FF6}} + \#\boldsymbol{w}_{\texttt{NewB6}} - \#\boldsymbol{w}_{\text{Inter}}} \tag{44}$$

$$\text{Precision}_{\texttt{FF6}} = \frac{\#\boldsymbol{w}_{\text{Inter}}}{\#\boldsymbol{w}_{\texttt{FF6}}} \tag{45}$$

$$\text{Precision}_{\texttt{NewB6}} = \frac{\#\boldsymbol{w}_{\text{Inter}}}{\#\boldsymbol{w}_{\texttt{NewB6}}} \tag{46}$$

$$F_1\text{-score} = \frac{2\#\boldsymbol{w}_{\text{Inter}}}{\#\boldsymbol{w}_{\texttt{FF6}} + \#\boldsymbol{w}_{\texttt{NewB6}}} \tag{47}$$

[9] Conventionally, one would label it as the coverage of FF6 by NewB6 ($\text{Recall}_{\texttt{NewB6}}$). Since the two benchmarks are treated symmetrically here, we instead name such an index as ($\text{Precision}_{\texttt{NewB6}}$).

Specifically, when using FF6 or NewB6 as the benchmark, the number of fake anomalies identified is 126 and 122, respectively. The intersection of $\boldsymbol{w}_{\texttt{FF6}}$ and $\boldsymbol{w}_{\texttt{NewB6}}$ contains 116 elements, indicating substantial overlap despite the two benchmarks sharing only Mkt. The magnitude of this overlap is also corroborated by a Jaccard index of 0.88 and an F_1 score of 0.94. Moreover, comparing the values of precision shows that this similarity is largely two-sided because the two are close.

4.2. Union of Fake Anomalies: What Remains?

Furthermore, we define $\boldsymbol{w}_{\text{Union}} = \boldsymbol{w}_{\texttt{FF6}} \cup \boldsymbol{w}_{\texttt{NewB6}}$, which contains 132 elements. The complementary set of $\boldsymbol{x}$ therefore has 21 elements; for ease of reference, we denote it by $\boldsymbol{x}^*$. Table 6 presents the detailed information on such anomalies. Since the FF6-based procedure yields 27 remaining anomalies, the factors classified as $\boldsymbol{x}$ under both benchmarks represent a large share of that set (78%), implying that the economic content we emphasize largely corresponds to what FF6 fails to span and that this content is robust to reasonable benchmark re-specification.

These 21 survivors coalesce into four economically distinct channels: (i) fundamental quality, (ii) management decisions, (iii) valuation level, and (iv) market behavior and patterns. Each channel highlights a source of risk that the benchmark factor model leaves unspanned or only partially absorbs.

The first channel, *fundamental quality*, integrates measures of a firm's underlying health, operating efficiency, and financial robustness. It encompasses 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). These indicators are cash-flow oriented and gauge resilience on the balance sheet, linking to cost structure and operating durability, and collectively identify firms that remain robust and sustain value creation. They are distinct from RMW, as they emphasize cash rather than book earnings and directly assess the reliability of accounting information. Relative to HML, they more effectively capture the persistence of

Table 6 Remaining Anomalies Across Benchmarks.

Note: This table reports the 21 anomalies that remain after taking the union of fake anomalies identified under the benchmarks FF6 and NewB6. For each anomaly, we provide its name, theme, concise description, and literature source.

Names of x	Theme	Description	Citation
cowc_gr1a	Accruals	Change in current operating working capital	Richardson et al. (2005)
oaccruals_at	Accruals	Operating accruals	Sloan (1996)
oaccruals_ni	Accruals	Percent operating accruals	Hafzalla et al. (2011)
noa_at	Debt issuance	Net operating assets	Hirshleifer et al. (2004)
cash_at	Low leverage	Cash-to-assets	Palazzo (2012)
netdebt_me	Low leverage	Net debt-to-price	Penman et al. (2007)
rd_sale	Low leverage	R&D-to-sales	Chan et al. (2001)
rd5_at	Low leverage	R&D capital-to-book assets	Li (2011)
sale_emp_gr1	Profit growth	Labor force efficiency	Abarbanell and Bushee (1998)
cop_at	Quality	Cash-based operating profits-to-book assets	Ball et al. (2016)
cop_atl1	Quality	Cash-based operating profits-to-lagged book assets	Ball et al. (2016)
gp_at	Quality	Gross profits-to-assets	Novy-Marx (2013)
qmj	Quality	Quality minus Junk: Composite	Asness et al. (2019)
qmj_safety	Quality	Quality minus Junk: Safety	Asness et al. (2019)
sale_bev	Quality	Assets turnover	Soliman (2008)
seas_6_10an	Seasonality	Years 6-10 lagged returns, annual	Heston and Sadka (2008)
seas_11_15an	Seasonality	Years 11-15 lagged returns, annual	Heston and Sadka (2008)
rd_me	Size	R&D-to-market	Chan et al. (2001)
rmax5_rvol_21d	Short-term reversal	Highest 5 days of return scaled by volatility	Asness et al. (2020)
fcf_me	Value	Free cash flow-to-price	Lakonishok et al. (1994)
div12m_me	Value	Dividend yield	Litzenberger and Ramaswamy (1979)

profitability and the provision of downside protection arising from superior operations and financial prudence. While the quality premium tends to be favored when liquidity is tight, the underlying mechanism does not arise solely from trading costs, so LIQ is at best a partial proxy for this exposure. As a result, anomalies in this channel are more likely to remain unspanned under reasonable benchmark choices.

The second channel, *management decisions*, centers on firms' strategic capital allocation and its implications for long-term value. It is captured by two distinct sets of measures: net operating assets (noa_at), which scale physical investment, and a suite of R&D intensity metrics (rd_sale, rd5_at, rd_me) that proxy for innovation effort. Unlike CMA, which emphasizes investment levels, and the IA, which captures physical investment intensity, this bundle focuses on the quality and efficiency of investment. Moreover, because R&D value is only imperfectly reflected on the balance sheet, these exposures do not fully overlap with the HML or the CMA. Consequently, anomalies in this channel retain a meaningful likelihood of remaining unspanned under both reasonable benchmark choices.

The third channel, *valuation level*, gauges price relative to fundamentals and is proxied by two shareholder cash-flow yields: free-cash-flow yield (`fcf_me`) and dividend yield (`div12m_me`). These measures quantify, at the prevailing market capitalization, cash that is actually distributed or feasibly distributable to shareholders. Because they are cash-based, they are less sensitive to noncash accounting charges (for example, depreciation) and more cleanly reflect the firm's value-creation capacity.

The fourth channel, *market behavior and patterns*, abstracts from firm fundamentals and focuses on pricing regularities driven by market microstructure and investor flows. It comprises short-horizon reversal (`rmax5_rvol_21d`), which reflects overreaction to extreme news followed by mean reversion, and long-lag seasonality (`seas_6_10an`, `seas_11_15an`), which is plausibly linked to calendar effects and institutional rebalancing cycles. These signals are not the mirror image of `MOM`: momentum captures intermediate-horizon trends, whereas short-horizon reversal reflects contrarian liquidity provision around temporary order-imbalance shocks, and seasonality reflects systematic timing in flows. These exposures are not a mirror image or simple re-expression of `BAB`. Their persistence indicates departures from frictionless, fully efficient markets; the associated risks, arising from trading frictions, behavioral biases, and flow dynamics, are absorbed only imperfectly by standard factor models.

Taken together, the final set of survivors sharpens the economic map of what the two benchmarks miss: fundamental quality, corporate policies and financing/investment choices, valuation, and market microstructure and flow-driven patterns. These dimensions can be considered robust to benchmark re-specification, deliver state-contingent insurance in high-marginal-utility states, and constitute disciplined candidates for augmenting the SDF beyond the benchmark menu.

4.3. Out-of-sample Evaluation

Essentially, the set formed by the final survivors together with FF6 (or NewB6) constitutes a candidate factor menu that may enter the SDF. Therefore, we have:

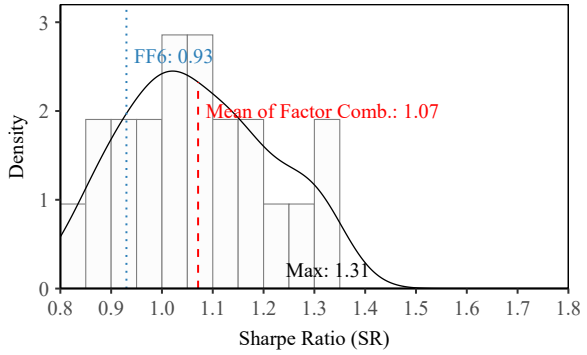
$$\boldsymbol{x}^* = \underbrace{\boldsymbol{\lambda}_{x^*}}_{(6+21)\times 1} + \boldsymbol{\epsilon}_{x^*,t}, \quad \boldsymbol{\epsilon}_{x^*,t} \sim \mathcal{N}_{27}(\mathbf{0}, \boldsymbol{\Omega}_{x^*}) \tag{48}$$

For simplicity, we present only the results combined with FF6. We first consider the OOS performance of tangency portfolios composed only of combinations of these potential risk factors. Following the process in Chib et al. (2024), for each factor combination, we recursively calculate the tangency portfolio for the following month; for each OOS month, we calculate the realized return of optimal risk factor tangency portfolios constructed in the previous month. We use these OOS realized returns to calculate the OOS Sharpe ratio for that set of factor combinations. We continue to use the first 25% of the sample to obtain prior information, and the final 120 months as the out-of-sample (OOS) evaluation window. Thus, the OOS period spans from January 2015 to December 2024, totaling 120 months.

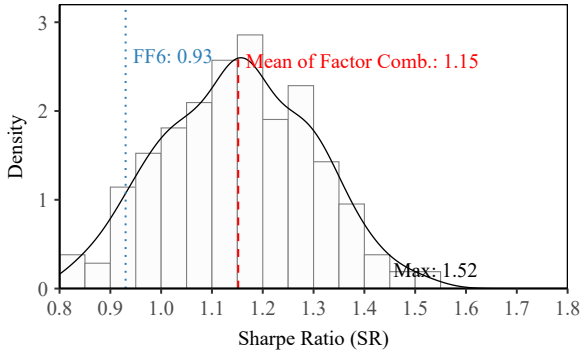
We compute the out-of-sample Sharpe ratio (OOS SR) for FF6 under this setup. We then enumerate several classes of factor combinations of $(\boldsymbol{x}, \texttt{FF6})$, compute the corresponding OOS SR statistics, and summarize their distributions. Figure 6 plots these distributions. As the number of added factors increases, both the mean and the maximum of the OOS SR distribution rise (the mean increases from 1.07 to 1.25) while the share of models with OOS SR below that of FF6 declines. Moreover, even when augmenting FF6 with a single $\boldsymbol{x}$ (Panel (a)), the mean OOS SR (1.07) exceeds that of FF6 alone (0.93). With five $\boldsymbol{x}$ added, the maximum attainable OOS SR reaches about 1.76.

Following Chib et al. (2024), we recover the marginal posteriors of the market-price vector $\boldsymbol{\Omega}_{x^*}^{-1}\boldsymbol{\lambda}_{x^*}$; Table 7 reports the summaries. On average, absolute market prices are largest for the quality-theme anomalies and the smallest for value-theme anomalies. Among the benchmark factors, Mkt, RMW, CMA, and MOM have 95% credible intervals that exclude zero, indicating strong evidence that they enter the SDF. Within the last-stage

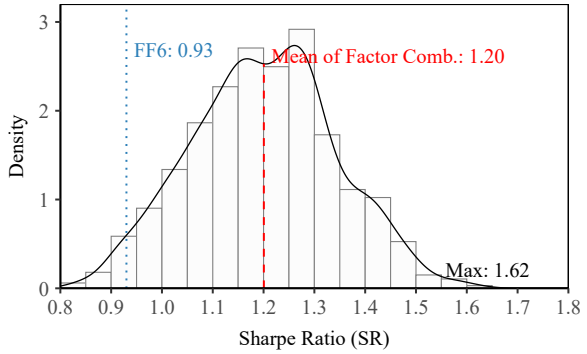
Figure 6 Distributions of Annualized Out-of-Sample Sharpe Ratios.
Note: This figure displays the distributions of the annualized OOS SR ratios for the factor combinations consisting of FF6 and last-stage anomalies based on 10,000 MCMC sample. We consider model classes formed by combining different numbers of $\boldsymbol{x}$ with the benchmark. Specifically, $(1\boldsymbol{x}, \texttt{FF6})$ denotes the 21 models obtained by augmenting FF6 with each single $\boldsymbol{x}$ in turn; for each model, we compute the corresponding OOS Sharpe ratio and then construct the reported statistic. The specifications $(2\boldsymbol{x}, \texttt{FF6})$ and higher orders are defined analogously. The red dashed line denotes the average OOS SR, and the blue dotted line shows the OOS SR obtained when using FF6 alone as the factor model.



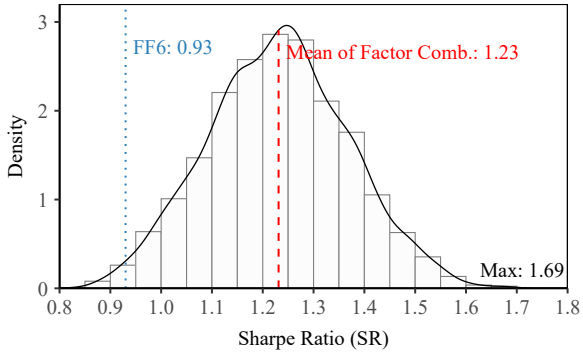
(a) $(1\boldsymbol{x}, \texttt{FF6})$



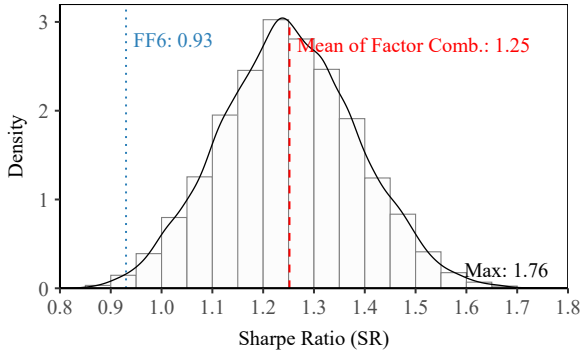
(b) $(2\boldsymbol{x}, \texttt{FF6})$



(c) $(3\boldsymbol{x}, \texttt{FF6})$



(d) $(4\boldsymbol{x}, \texttt{FF6})$



(e) $(5\boldsymbol{x}, \texttt{FF6})$

anomalies, seas_6_10an, seas_11_15an, rd_me, and rmax5_rvol_21d have 95% credible intervals that exclude zero. This evidence suggests that, relative to the remaining factors, these eight are more likely to be part of the true SDF.

Table 7 Posterior Statistics of the Market Prices of Factor Risks of FF6 and Last-stage Anomalies.

Note: This table reports the posterior statistics, that is, posterior mean, posterior standard deviation, posterior median, 2.5% and 97.5% quantiles for the market prices of factor risks $\boldsymbol{\Omega}_{x^*}^{-1}\boldsymbol{\lambda}_{x^*}$.

Name	Theme	Post. mean	Post. std.	Post. median	2.5% qtile	97.5% qtile
Panel A. Benchmark						
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**
Panel B. Last-stage Anomalies						
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

5. Conclusion

Motivated by the roots of the "replication crisis," we emphasize a simple but often overlooked point: in an environment of anomaly proliferation and multiple testing, establishing that a signal is not spanned would require knowledge of the true SDF, an object that is essentially unattainable in practice. In contrast, establishing that a signal is spanned only requires exhibiting a factor set that linearly spans it (i.e., zero-intercept pricing). Building on this shift in the burden of proof, we develop a framework centered on Bayesian model partitioning and marginal likelihood, coupled with EFDR control and a stepwise scan. We place benchmark factors and candidate anomalies in a common model space, enumerate all partitions into "unspanned/spanned," and compute for each anomaly its posterior probability of being spanned. We then apply an EFDR threshold and iteratively filter out anomalies that are statistically spanned. In this way, the question moves from "*Is this anomaly an anomaly*" to "*Can this anomaly be spanned?*"

Methodologically, our novel approach tackles two challenges at once. First, model uncertainty. Rather than a binary comparison of the FF benchmark with and without intercept, we integrate over all spanned/unspanned partitions and summarize the evidence via posterior model probabilities. Second, multiplicity. We control EFDR and select the set of spanned anomalies. Crucially, the workflow is deliberately asymmetric toward establishing span, conservative with respect to declaring true anomalies. As a result, failure to be spanned by the current benchmark is not mistaken for evidence of anomaly, and conclusions about the replication crisis do not hinge on a fragile optimal-benchmark assumption.

Using the 153 U.S. factors (Jensen et al., 2023) from 1985 to 2024, our framework delivers clear evidence. With `FF6` as the benchmark, the stepwise scan proceeds as follows: steps 1, 2, 4, and 5 filter 21, 87, 5, and 13 fake anomalies, respectively (step 3 selects none), stabilizing at 126 spanned and 27 unspanned anomalies. These findings are not mere statistical artifacts: under basic frequentist and Bayesian tests, `FF6` plus 1 to 5 of the $\boldsymbol{x}$ explain all 126 fake anomalies. On the investment side, portfolios consisting of

FF6 plus 5 or 6 of the last-stage anomalies deliver in-sample and out-of-sample tangency-portfolio Sharpe ratios and cross-sectional fit for industry, bi-sort, and P-Tree test assets that systematically exceed those of fake anomaly sets of the same size.

We find that the cross-benchmark (FF6 and NewB6) union of spanned anomalies equals 132, leaving a residual of 21 anomalies in the unspanned set. The intersection of the two benchmark results closely matches the FF6-based set, indicating that the economic content we isolate largely corresponds to what FF6 fails to span and that the finding is robust to reasonable benchmark re-specification. Furthermore, we find eight factors, four from FF6 and four from the 21 anomalies, whose credible intervals of the posterior distributions of the market prices of risk are strictly positive (excluding 0), indicating that these eight factors are the strongest candidates to enter the true SDF.

Moreover, the 21 survivors cohere along four economically linked yet only partially overlapping channels. The stability and economic coherence of these 21 survivors indicate non-redundant SDF exposures. These deliver state-contingent insurance in high-marginal-utility states and generate verifiable incremental cross-sectional pricing power. As such, they constitute candidates for augmenting the SDF beyond the standard benchmark menu.

In summary, the paper makes three contributions to the replication-crisis literature. First, we transform the untestable claim of anomaly existence into a testable claim about spanning, and supply reproducible decision rules based on posterior model probabilities and EFDR. Second, we introduce an implementable stepwise scan that enables systematic screening in a high-dimensional anomaly zoo. Third, we provide broad-based and robust evidence that most reported signals are spanned in a benchmark plus a few anomalies setting, while the small set of stable survivors exhibits a coherent economic structure and interpretable risk-friction channels.

References

Abarbanell, J. S. and B. J. Bushee (1998). Abnormal returns to a fundamental analysis strategy. *The Accounting Review 73*(1), 19–45.

Asness, C. and A. Frazzini (2013). The devil in hml's details. *Journal of Portfolio Management 39*(4), 49–68.

Asness, C., A. Frazzini, N. J. Gormsen, and L. H. Pedersen (2020). Betting against correlation: Testing theories of the low-risk effect. *Journal of Financial Economics 135*(3), 629–652.

Asness, C. S., A. Frazzini, and L. H. Pedersen (2019). Quality minus junk. *Review of Accounting Studies 24*(1), 34–112.

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. Nikolaev (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics 121*(1), 28–45.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Chan, L. K., J. Lakonishok, and T. Sougiannis (2001). The stock market valuation of research and development expenditures. *Journal of Finance 56*(6), 2431–2456.

Chen, A. Y. (2025). Do t-statistic hurdles need to be raised? *Management Science 71*(7), 5830–5848.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association 90*(432), 1313–1321.

Chib, S. and X. Zeng (2020). Which factors are risk factors in asset pricing? A model scan framework. *Journal of Business & Economic Statistics 38*, 771–783.

Chib, S., X. Zeng, and L. Zhao (2020). On comparing asset pricing models. *Journal of Finance 75*(1), 551–577.

Chib, S., L. Zhao, and G. Zhou (2024). Winners from winners: A tale of risk factors. *Management Science 70*(1), 396–414.

Chordia, T., A. Goyal, and A. Saretto (2020). Anomalies and false rejections. *Review of Financial Studies 33*(5), 2134–2179.

Cong, L. W., G. Feng, J. He, and X. He (2025). Growing the efficient frontier on panel trees. *Journal of Financial Economics 167*, 104024.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and K. R. French (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies 29*(1), 69–103.

Frazzini, A. and L. H. Pedersen (2014). Betting Against Beta. *Journal of Financial Economics 111*(1), 1–25.

Hafzalla, N., R. Lundholm, and E. Matthew Van Winkle (2011). Percent accruals. *The Accounting Review 86*(1), 209–236.

Harvey, C. R. (2017). Presidential address: The scientific outlook in financial economics. *Journal of Finance 72*(4), 1399–1440.

Harvey, C. R. and Y. Liu (2020). False (and missed) discoveries in financial economics. *Journal of Finance 75*(5), 2503–2553.

Harvey, C. R., Y. Liu, and A. Saretto (2020). An evaluation of alternative multiple testing methods for finance applications. *Review of Asset Pricing Studies 10*(2), 199–248.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

Heston, S. L. and R. Sadka (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics 87*(2), 418–445.

Hirshleifer, D., K. Hou, S. H. Teoh, and Y. Zhang (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics 38*, 297–331.

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *Review of Financial Studies 28*(3), 650–705.

Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *Review of Financial Studies 33*(5), 2019–2133.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2023). Is there a replication crisis in finance? *Journal of Finance 78*(5), 2465–2518.

Lakonishok, J., A. Shleifer, and R. W. Vishny (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance 49*(5), 1541–1578.

Lehmann, B. N. and D. M. Modest (1987). Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *Journal of Finance 42*(2), 233–265.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics 96*(2), 175–194.

Li, F. (2011). Earnings quality based on corporate investment decisions. *Journal of Accounting Research 49*(3), 721–752.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics 47*, 13–37.

Litzenberger, R. H. and K. Ramaswamy (1979). The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics 7*(2), 163–195.

Müller, P., G. Parmigiani, and K. Rice (2007). FDR and Bayesian multiple comparisons rules. In *Bayesian Statistics 8*, pp. 359–380. Oxford University Press.

Newton, M. A., A. Noueiry, D. Sarkar, and P. Ahlquist (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics 5*(2), 155–176.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics 108*(1), 1–28.

Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics 104*(1), 162–185.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Penman, S. H., S. A. Richardson, and I. Tuna (2007). The book-to-price effect in stock returns: Accounting for leverage. *Journal of Accounting Research 45*(2), 427–467.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics 39*(3), 437–485.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*(3), 425–442.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review 71*(3), 289–315.

Soliman, M. T. (2008). The use of DuPont analysis by market participants. *The Accounting Review 83*(3), 823–853.

Yan, X. and L. Zheng (2017). Fundamental analysis and the cross-section of stock returns: A data-mining approach. *Review of Financial Studies 30*(4), 1382–1423.

Appendix

Table A.1 Candidate Splits for Factor Decomposition (Benchmark: `FF6`; one at a time). *Shading key:* gray (shaded) rows are models where the candidate anomaly a_i is *spanned* (i.e., $a_i \in \boldsymbol{w}$); unshaded rows are models where it is *unspanned* (i.e., $a_i \in \boldsymbol{x}$).

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
$\mathbb{M}_{10}^{(i)}$	Mkt, RMW	SMB, HML, CMA, MOM, a_i
$\mathbb{M}_{11}^{(i)}$	Mkt, CMA	SMB, HML, RMW, MOM, a_i
$\mathbb{M}_{12}^{(i)}$	Mkt, MOM	SMB, HML, RMW, CMA, a_i
$\mathbb{M}_{13}^{(i)}$	Mkt, a_i	SMB, HML, RMW, CMA, MOM
$\mathbb{M}_{14}^{(i)}$	SMB, HML	Mkt, RMW, CMA, MOM, a_i
$\mathbb{M}_{15}^{(i)}$	SMB, RMW	Mkt, HML, CMA, MOM, a_i
$\mathbb{M}_{16}^{(i)}$	SMB, CMA	Mkt, HML, RMW, MOM, a_i
$\mathbb{M}_{17}^{(i)}$	SMB, MOM	Mkt, HML, RMW, CMA, a_i
$\mathbb{M}_{18}^{(i)}$	SMB, a_i	Mkt, HML, RMW, CMA, MOM
$\mathbb{M}_{19}^{(i)}$	HML, RMW	Mkt, SMB, CMA, MOM, a_i
$\mathbb{M}_{20}^{(i)}$	HML, CMA	Mkt, SMB, RMW, MOM, a_i
$\mathbb{M}_{21}^{(i)}$	HML, MOM	Mkt, SMB, RMW, CMA, a_i
$\mathbb{M}_{22}^{(i)}$	HML, a_i	Mkt, SMB, RMW, CMA, MOM
$\mathbb{M}_{23}^{(i)}$	RMW, CMA	Mkt, SMB, HML, MOM, a_i
$\mathbb{M}_{24}^{(i)}$	RMW, MOM	Mkt, SMB, HML, CMA, a_i
$\mathbb{M}_{25}^{(i)}$	RMW, a_i	Mkt, SMB, HML, CMA, MOM
$\mathbb{M}_{26}^{(i)}$	CMA, MOM	Mkt, SMB, HML, RMW, a_i
$\mathbb{M}_{27}^{(i)}$	CMA, a_i	Mkt, SMB, HML, RMW, MOM
$\mathbb{M}_{28}^{(i)}$	MOM, a_i	Mkt, SMB, HML, RMW, CMA
$\mathbb{M}_{29}^{(i)}$	Mkt, SMB, HML	RMW, CMA, MOM, a_i
$\mathbb{M}_{30}^{(i)}$	Mkt, SMB, RMW	HML, CMA, MOM, a_i
$\mathbb{M}_{31}^{(i)}$	Mkt, SMB, CMA	HML, RMW, MOM, a_i
$\mathbb{M}_{32}^{(i)}$	Mkt, SMB, MOM	HML, RMW, CMA, a_i
$\mathbb{M}_{33}^{(i)}$	Mkt, SMB, a_i	HML, RMW, CMA, MOM
$\mathbb{M}_{34}^{(i)}$	Mkt, HML, RMW	SMB, CMA, MOM, a_i
$\mathbb{M}_{35}^{(i)}$	Mkt, HML, CMA	SMB, RMW, MOM, a_i
$\mathbb{M}_{36}^{(i)}$	Mkt, HML, MOM	SMB, RMW, CMA, a_i
$\mathbb{M}_{37}^{(i)}$	Mkt, HML, a_i	SMB, RMW, CMA, MOM
$\mathbb{M}_{38}^{(i)}$	Mkt, RMW, CMA	SMB, HML, MOM, a_i

continued on next page

Table A.1 (continued)

Model	$\boldsymbol{x}$	$\boldsymbol{w}$
$\mathbb{M}_{39}^{(i)}$	Mkt, RMW, MOM	SMB, HML, CMA, a_i
$\mathbb{M}_{40}^{(i)}$	Mkt, RMW, a_i	SMB, HML, CMA, MOM
$\mathbb{M}_{41}^{(i)}$	Mkt, CMA, MOM	SMB, HML, RMW, a_i
$\mathbb{M}_{42}^{(i)}$	Mkt, CMA, a_i	SMB, HML, RMW, MOM
$\mathbb{M}_{43}^{(i)}$	Mkt, MOM, a_i	SMB, HML, RMW, CMA
$\mathbb{M}_{44}^{(i)}$	SMB, HML, RMW	Mkt, CMA, MOM, a_i
$\mathbb{M}_{45}^{(i)}$	SMB, HML, CMA	Mkt, RMW, MOM, a_i
$\mathbb{M}_{46}^{(i)}$	SMB, HML, MOM	Mkt, RMW, CMA, a_i
$\mathbb{M}_{47}^{(i)}$	SMB, HML, a_i	Mkt, RMW, CMA, MOM
$\mathbb{M}_{48}^{(i)}$	SMB, RMW, CMA	Mkt, HML, MOM, a_i
$\mathbb{M}_{49}^{(i)}$	SMB, RMW, MOM	Mkt, HML, CMA, a_i
$\mathbb{M}_{50}^{(i)}$	SMB, RMW, a_i	Mkt, HML, CMA, MOM
$\mathbb{M}_{51}^{(i)}$	SMB, CMA, MOM	Mkt, HML, RMW, a_i
$\mathbb{M}_{52}^{(i)}$	SMB, CMA, a_i	Mkt, HML, RMW, MOM
$\mathbb{M}_{53}^{(i)}$	SMB, MOM, a_i	Mkt, HML, RMW, CMA
$\mathbb{M}_{54}^{(i)}$	HML, RMW, CMA	Mkt, SMB, MOM, a_i
$\mathbb{M}_{55}^{(i)}$	HML, RMW, MOM	Mkt, SMB, CMA, a_i
$\mathbb{M}_{56}^{(i)}$	HML, RMW, a_i	Mkt, SMB, CMA, MOM
$\mathbb{M}_{57}^{(i)}$	HML, CMA, MOM	Mkt, SMB, RMW, a_i
$\mathbb{M}_{58}^{(i)}$	HML, CMA, a_i	Mkt, SMB, RMW, MOM
$\mathbb{M}_{59}^{(i)}$	HML, MOM, a_i	Mkt, SMB, RMW, CMA
$\mathbb{M}_{60}^{(i)}$	RMW, CMA, MOM	Mkt, SMB, HML, a_i
$\mathbb{M}_{61}^{(i)}$	RMW, CMA, a_i	Mkt, SMB, HML, MOM
$\mathbb{M}_{62}^{(i)}$	RMW, MOM, a_i	Mkt, SMB, HML, CMA
$\mathbb{M}_{63}^{(i)}$	CMA, MOM, a_i	Mkt, SMB, HML, RMW
$\mathbb{M}_{64}^{(i)}$	Mkt, SMB, HML, RMW	CMA, MOM, a_i
$\mathbb{M}_{65}^{(i)}$	Mkt, SMB, HML, CMA	RMW, MOM, a_i
$\mathbb{M}_{66}^{(i)}$	Mkt, SMB, HML, MOM	RMW, CMA, a_i
$\mathbb{M}_{67}^{(i)}$	Mkt, SMB, HML, a_i	RMW, CMA, MOM
$\mathbb{M}_{68}^{(i)}$	Mkt, SMB, RMW, CMA	HML, MOM, a_i
$\mathbb{M}_{69}^{(i)}$	Mkt, SMB, RMW, MOM	HML, CMA, a_i
$\mathbb{M}_{70}^{(i)}$	Mkt, SMB, RMW, a_i	HML, CMA, MOM
$\mathbb{M}_{71}^{(i)}$	Mkt, SMB, CMA, MOM	HML, RMW, a_i
$\mathbb{M}_{72}^{(i)}$	Mkt, SMB, CMA, a_i	HML, RMW, MOM
$\mathbb{M}_{73}^{(i)}$	Mkt, SMB, MOM, a_i	HML, RMW, CMA
$\mathbb{M}_{74}^{(i)}$	Mkt, HML, RMW, CMA	SMB, MOM, a_i
$\mathbb{M}_{75}^{(i)}$	Mkt, HML, RMW, MOM	SMB, CMA, a_i
$\mathbb{M}_{76}^{(i)}$	Mkt, HML, RMW, a_i	SMB, CMA, MOM
$\mathbb{M}_{77}^{(i)}$	Mkt, HML, CMA, MOM	SMB, RMW, a_i
$\mathbb{M}_{78}^{(i)}$	Mkt, HML, CMA, a_i	SMB, RMW, MOM
$\mathbb{M}_{79}^{(i)}$	Mkt, HML, MOM, a_i	SMB, RMW, CMA
$\mathbb{M}_{80}^{(i)}$	Mkt, RMW, CMA, MOM	SMB, HML, a_i
$\mathbb{M}_{81}^{(i)}$	Mkt, RMW, CMA, a_i	SMB, HML, MOM
$\mathbb{M}_{82}^{(i)}$	Mkt, RMW, MOM, a_i	SMB, HML, CMA
$\mathbb{M}_{83}^{(i)}$	Mkt, CMA, MOM, a_i	SMB, HML, RMW

continued on next page

Table A.1 (continued)

Model	$\boldsymbol{x}$	$\boldsymbol{w}$
$\mathbb{M}_{84}^{(i)}$	SMB, HML, RMW, CMA	Mkt, MOM, a_i
$\mathbb{M}_{85}^{(i)}$	SMB, HML, RMW, MOM	Mkt, CMA, a_i
$\mathbb{M}_{86}^{(i)}$	SMB, HML, RMW, a_i	Mkt, CMA, MOM
$\mathbb{M}_{87}^{(i)}$	SMB, HML, CMA, MOM	Mkt, RMW, a_i
$\mathbb{M}_{88}^{(i)}$	SMB, HML, CMA, a_i	Mkt, RMW, MOM
$\mathbb{M}_{89}^{(i)}$	SMB, HML, MOM, a_i	Mkt, RMW, CMA
$\mathbb{M}_{90}^{(i)}$	SMB, RMW, CMA, MOM	Mkt, HML, a_i
$\mathbb{M}_{91}^{(i)}$	SMB, RMW, CMA, a_i	Mkt, HML, MOM
$\mathbb{M}_{92}^{(i)}$	SMB, RMW, MOM, a_i	Mkt, HML, CMA
$\mathbb{M}_{93}^{(i)}$	SMB, CMA, MOM, a_i	Mkt, HML, RMW
$\mathbb{M}_{94}^{(i)}$	HML, RMW, CMA, MOM	Mkt, SMB, a_i
$\mathbb{M}_{95}^{(i)}$	HML, RMW, CMA, a_i	Mkt, SMB, MOM
$\mathbb{M}_{96}^{(i)}$	HML, RMW, MOM, a_i	Mkt, SMB, CMA
$\mathbb{M}_{97}^{(i)}$	HML, CMA, MOM, a_i	Mkt, SMB, RMW
$\mathbb{M}_{98}^{(i)}$	RMW, CMA, MOM, a_i	Mkt, SMB, HML
$\mathbb{M}_{99}^{(i)}$	Mkt, SMB, HML, RMW, CMA	MOM, a_i
$\mathbb{M}_{100}^{(i)}$	Mkt, SMB, HML, RMW, MOM	CMA, a_i
$\mathbb{M}_{101}^{(i)}$	Mkt, SMB, HML, RMW, a_i	CMA, MOM
$\mathbb{M}_{102}^{(i)}$	Mkt, SMB, HML, CMA, MOM	RMW, a_i
$\mathbb{M}_{103}^{(i)}$	Mkt, SMB, HML, CMA, a_i	RMW, MOM
$\mathbb{M}_{104}^{(i)}$	Mkt, SMB, HML, MOM, a_i	RMW, CMA
$\mathbb{M}_{105}^{(i)}$	Mkt, SMB, RMW, CMA, MOM	HML, a_i
$\mathbb{M}_{106}^{(i)}$	Mkt, SMB, RMW, CMA, a_i	HML, MOM
$\mathbb{M}_{107}^{(i)}$	Mkt, SMB, RMW, MOM, a_i	HML, CMA
$\mathbb{M}_{108}^{(i)}$	Mkt, SMB, CMA, MOM, a_i	HML, RMW
$\mathbb{M}_{109}^{(i)}$	Mkt, HML, RMW, CMA, MOM	SMB, a_i
$\mathbb{M}_{110}^{(i)}$	Mkt, HML, RMW, CMA, a_i	SMB, MOM
$\mathbb{M}_{111}^{(i)}$	Mkt, HML, RMW, MOM, a_i	SMB, CMA
$\mathbb{M}_{112}^{(i)}$	Mkt, HML, CMA, MOM, a_i	SMB, RMW
$\mathbb{M}_{113}^{(i)}$	Mkt, RMW, CMA, MOM, a_i	SMB, HML
$\mathbb{M}_{114}^{(i)}$	SMB, HML, RMW, CMA, MOM	Mkt, a_i
$\mathbb{M}_{115}^{(i)}$	SMB, HML, RMW, CMA, a_i	Mkt, MOM
$\mathbb{M}_{116}^{(i)}$	SMB, HML, RMW, MOM, a_i	Mkt, CMA
$\mathbb{M}_{117}^{(i)}$	SMB, HML, CMA, MOM, a_i	Mkt, RMW
$\mathbb{M}_{118}^{(i)}$	SMB, RMW, CMA, MOM, a_i	Mkt, HML
$\mathbb{M}_{119}^{(i)}$	HML, RMW, CMA, MOM, a_i	Mkt, SMB
$\mathbb{M}_{120}^{(i)}$	Mkt, SMB, HML, RMW, CMA, MOM	a_i
$\mathbb{M}_{121}^{(i)}$	Mkt, SMB, HML, RMW, CMA, a_i	MOM
$\mathbb{M}_{122}^{(i)}$	Mkt, SMB, HML, RMW, MOM, a_i	CMA
$\mathbb{M}_{123}^{(i)}$	Mkt, SMB, HML, CMA, MOM, a_i	RMW
$\mathbb{M}_{124}^{(i)}$	Mkt, SMB, RMW, CMA, MOM, a_i	HML
$\mathbb{M}_{125}^{(i)}$	Mkt, HML, RMW, CMA, MOM, a_i	SMB
$\mathbb{M}_{126}^{(i)}$	SMB, HML, RMW, CMA, MOM, a_i	Mkt
$\mathbb{M}_{127}^{(i)}$	Mkt, SMB, HML, RMW, CMA, MOM, a_i	$\varnothing$

Table A.2 Candidate Splits for Factor Decomposition (Benchmark: `FF6`; two at a time). *Shading key:* gray (shaded) rows are models where the pair (b_i, b_k) *is* jointly spanned ($\{b_i, b_k\} \subseteq \boldsymbol{w}$); unshaded rows are models where they are *jointly unspanned* ($\{b_i, b_k\} \subseteq \boldsymbol{x}$). Mixed configurations are excluded.

Model	$\boldsymbol{x}$	$\boldsymbol{w}$
$\mathbb{M}_1^{(i,k)}$	Mkt	SMB, HML, RMW, CMA, MOM, b_i, b_k
$\mathbb{M}_2^{(i,k)}$	SMB	Mkt, HML, RMW, CMA, MOM, b_i, b_k
$\mathbb{M}_3^{(i,k)}$	HML	Mkt, SMB, RMW, CMA, MOM, b_i, b_k
$\mathbb{M}_4^{(i,k)}$	RMW	Mkt, SMB, HML, CMA, MOM, b_i, b_k
$\mathbb{M}_5^{(i,k)}$	CMA	Mkt, SMB, HML, RMW, MOM, b_i, b_k
$\mathbb{M}_6^{(i,k)}$	MOM	Mkt, SMB, HML, RMW, CMA, b_i, b_k
$\mathbb{M}_7^{(i,k)}$	b_i, b_k	Mkt, SMB, HML, RMW, CMA, MOM
$\mathbb{M}_8^{(i,k)}$	Mkt, SMB	HML, RMW, CMA, MOM, b_i, b_k
$\mathbb{M}_9^{(i,k)}$	Mkt, HML	SMB, RMW, CMA, MOM, b_i, b_k
$\mathbb{M}_{10}^{(i,k)}$	Mkt, RMW	SMB, HML, CMA, MOM, b_i, b_k
$\mathbb{M}_{11}^{(i,k)}$	Mkt, CMA	SMB, HML, RMW, MOM, b_i, b_k
$\mathbb{M}_{12}^{(i,k)}$	Mkt, MOM	SMB, HML, RMW, CMA, b_i, b_k
$\mathbb{M}_{13}^{(i,k)}$	Mkt, b_i, b_k	SMB, HML, RMW, CMA, MOM
$\mathbb{M}_{14}^{(i,k)}$	SMB, HML	Mkt, RMW, CMA, MOM, b_i, b_k
$\mathbb{M}_{15}^{(i,k)}$	SMB, RMW	Mkt, HML, CMA, MOM, b_i, b_k
$\mathbb{M}_{16}^{(i,k)}$	SMB, CMA	Mkt, HML, RMW, MOM, b_i, b_k
$\mathbb{M}_{17}^{(i,k)}$	SMB, MOM	Mkt, HML, RMW, CMA, b_i, b_k
$\mathbb{M}_{18}^{(i,k)}$	SMB, b_i, b_k	Mkt, HML, RMW, CMA, MOM
$\mathbb{M}_{19}^{(i,k)}$	HML, RMW	Mkt, SMB, CMA, MOM, b_i, b_k
$\mathbb{M}_{20}^{(i,k)}$	HML, CMA	Mkt, SMB, RMW, MOM, b_i, b_k
$\mathbb{M}_{21}^{(i,k)}$	HML, MOM	Mkt, SMB, RMW, CMA, b_i, b_k
$\mathbb{M}_{22}^{(i,k)}$	HML, b_i, b_k	Mkt, SMB, RMW, CMA, MOM
$\mathbb{M}_{23}^{(i,k)}$	RMW, CMA	Mkt, SMB, HML, MOM, b_i, b_k
$\mathbb{M}_{24}^{(i,k)}$	RMW, MOM	Mkt, SMB, HML, CMA, b_i, b_k
$\mathbb{M}_{25}^{(i,k)}$	RMW, b_i, b_k	Mkt, SMB, HML, CMA, MOM
$\mathbb{M}_{26}^{(i,k)}$	CMA, MOM	Mkt, SMB, HML, RMW, b_i, b_k
$\mathbb{M}_{27}^{(i,k)}$	CMA, b_i, b_k	Mkt, SMB, HML, RMW, MOM
$\mathbb{M}_{28}^{(i,k)}$	MOM, b_i, b_k	Mkt, SMB, HML, RMW, CMA
$\mathbb{M}_{29}^{(i,k)}$	Mkt, SMB, HML	RMW, CMA, MOM, b_i, b_k
$\mathbb{M}_{30}^{(i,k)}$	Mkt, SMB, RMW	HML, CMA, MOM, b_i, b_k
$\mathbb{M}_{31}^{(i,k)}$	Mkt, SMB, CMA	HML, RMW, MOM, b_i, b_k
$\mathbb{M}_{32}^{(i,k)}$	Mkt, SMB, MOM	HML, RMW, CMA, b_i, b_k
$\mathbb{M}_{33}^{(i,k)}$	Mkt, SMB, b_i, b_k	HML, RMW, CMA, MOM
$\mathbb{M}_{34}^{(i,k)}$	Mkt, HML, RMW	SMB, CMA, MOM, b_i, b_k
$\mathbb{M}_{35}^{(i,k)}$	Mkt, HML, CMA	SMB, RMW, MOM, b_i, b_k
$\mathbb{M}_{36}^{(i,k)}$	Mkt, HML, MOM	SMB, RMW, CMA, b_i, b_k
$\mathbb{M}_{37}^{(i,k)}$	Mkt, HML, b_i, b_k	SMB, RMW, CMA, MOM
$\mathbb{M}_{38}^{(i,k)}$	Mkt, RMW, CMA	SMB, HML, MOM, b_i, b_k
$\mathbb{M}_{39}^{(i,k)}$	Mkt, RMW, MOM	SMB, HML, CMA, b_i, b_k
$\mathbb{M}_{40}^{(i,k)}$	Mkt, RMW, b_i, b_k	SMB, HML, CMA, MOM
$\mathbb{M}_{41}^{(i,k)}$	Mkt, CMA, MOM	SMB, HML, RMW, b_i, b_k

continued on next page

Table A.2 (continued)

Model	$\boldsymbol{x}$	$\boldsymbol{w}$
$\mathbb{M}_{42}^{(i,k)}$	Mkt, CMA, b_i, b_k	SMB, HML, RMW, MOM
$\mathbb{M}_{43}^{(i,k)}$	Mkt, MOM, b_i, b_k	SMB, HML, RMW, CMA
$\mathbb{M}_{44}^{(i,k)}$	SMB, HML, RMW	Mkt, CMA, MOM, b_i, b_k
$\mathbb{M}_{45}^{(i,k)}$	SMB, HML, CMA	Mkt, RMW, MOM, b_i, b_k
$\mathbb{M}_{46}^{(i,k)}$	SMB, HML, MOM	Mkt, RMW, CMA, b_i, b_k
$\mathbb{M}_{47}^{(i,k)}$	SMB, HML, b_i, b_k	Mkt, RMW, CMA, MOM
$\mathbb{M}_{48}^{(i,k)}$	SMB, RMW, CMA	Mkt, HML, MOM, b_i, b_k
$\mathbb{M}_{49}^{(i,k)}$	SMB, RMW, MOM	Mkt, HML, CMA, b_i, b_k
$\mathbb{M}_{50}^{(i,k)}$	SMB, RMW, b_i, b_k	Mkt, HML, CMA, MOM
$\mathbb{M}_{51}^{(i,k)}$	SMB, CMA, MOM	Mkt, HML, RMW, b_i, b_k
$\mathbb{M}_{52}^{(i,k)}$	SMB, CMA, b_i, b_k	Mkt, HML, RMW, MOM
$\mathbb{M}_{53}^{(i,k)}$	SMB, MOM, b_i, b_k	Mkt, HML, RMW, CMA
$\mathbb{M}_{54}^{(i,k)}$	HML, RMW, CMA	Mkt, SMB, MOM, b_i, b_k
$\mathbb{M}_{55}^{(i,k)}$	HML, RMW, MOM	Mkt, SMB, CMA, b_i, b_k
$\mathbb{M}_{56}^{(i,k)}$	HML, RMW, b_i, b_k	Mkt, SMB, CMA, MOM
$\mathbb{M}_{57}^{(i,k)}$	HML, CMA, MOM	Mkt, SMB, RMW, b_i, b_k
$\mathbb{M}_{58}^{(i,k)}$	HML, CMA, b_i, b_k	Mkt, SMB, RMW, MOM
$\mathbb{M}_{59}^{(i,k)}$	HML, MOM, b_i, b_k	Mkt, SMB, RMW, CMA
$\mathbb{M}_{60}^{(i,k)}$	RMW, CMA, MOM	Mkt, SMB, HML, b_i, b_k
$\mathbb{M}_{61}^{(i,k)}$	RMW, CMA, b_i, b_k	Mkt, SMB, HML, MOM
$\mathbb{M}_{62}^{(i,k)}$	RMW, MOM, b_i, b_k	Mkt, SMB, HML, CMA
$\mathbb{M}_{63}^{(i,k)}$	CMA, MOM, b_i, b_k	Mkt, SMB, HML, RMW
$\mathbb{M}_{64}^{(i,k)}$	Mkt, SMB, HML, RMW	CMA, MOM, b_i, b_k
$\mathbb{M}_{65}^{(i,k)}$	Mkt, SMB, HML, CMA	RMW, MOM, b_i, b_k
$\mathbb{M}_{66}^{(i,k)}$	Mkt, SMB, HML, MOM	RMW, CMA, b_i, b_k
$\mathbb{M}_{67}^{(i,k)}$	Mkt, SMB, HML, b_i, b_k	RMW, CMA, MOM
$\mathbb{M}_{68}^{(i,k)}$	Mkt, SMB, RMW, CMA	HML, MOM, b_i, b_k
$\mathbb{M}_{69}^{(i,k)}$	Mkt, SMB, RMW, MOM	HML, CMA, b_i, b_k
$\mathbb{M}_{70}^{(i,k)}$	Mkt, SMB, RMW, b_i, b_k	HML, CMA, MOM
$\mathbb{M}_{71}^{(i,k)}$	Mkt, SMB, CMA, MOM	HML, RMW, b_i, b_k
$\mathbb{M}_{72}^{(i,k)}$	Mkt, SMB, CMA, b_i, b_k	HML, RMW, MOM
$\mathbb{M}_{73}^{(i,k)}$	Mkt, SMB, MOM, b_i, b_k	HML, RMW, CMA
$\mathbb{M}_{74}^{(i,k)}$	Mkt, HML, RMW, CMA	SMB, MOM, b_i, b_k
$\mathbb{M}_{75}^{(i,k)}$	Mkt, HML, RMW, MOM	SMB, CMA, b_i, b_k
$\mathbb{M}_{76}^{(i,k)}$	Mkt, HML, RMW, b_i, b_k	SMB, CMA, MOM
$\mathbb{M}_{77}^{(i,k)}$	Mkt, HML, CMA, MOM	SMB, RMW, b_i, b_k
$\mathbb{M}_{78}^{(i,k)}$	Mkt, HML, CMA, b_i, b_k	SMB, RMW, MOM
$\mathbb{M}_{79}^{(i,k)}$	Mkt, HML, MOM, b_i, b_k	SMB, RMW, CMA
$\mathbb{M}_{80}^{(i,k)}$	Mkt, RMW, CMA, MOM	SMB, HML, b_i, b_k
$\mathbb{M}_{81}^{(i,k)}$	Mkt, RMW, CMA, b_i, b_k	SMB, HML, MOM
$\mathbb{M}_{82}^{(i,k)}$	Mkt, RMW, MOM, b_i, b_k	SMB, HML, CMA
$\mathbb{M}_{83}^{(i,k)}$	Mkt, CMA, MOM, b_i, b_k	SMB, HML, RMW
$\mathbb{M}_{84}^{(i,k)}$	SMB, HML, RMW, CMA	Mkt, MOM, b_i, b_k
$\mathbb{M}_{85}^{(i,k)}$	SMB, HML, RMW, MOM	Mkt, CMA, b_i, b_k
$\mathbb{M}_{86}^{(i,k)}$	SMB, HML, RMW, b_i, b_k	Mkt, CMA, MOM

continued on next page

Table A.2 (continued)

Model	$\boldsymbol{x}$	$\boldsymbol{w}$
$\mathbb{M}_{87}^{(i,k)}$	SMB, HML, CMA, MOM	Mkt, RMW, b_i, b_k
$\mathbb{M}_{88}^{(i,k)}$	SMB, HML, CMA, b_i, b_k	Mkt, RMW, MOM
$\mathbb{M}_{89}^{(i,k)}$	SMB, HML, MOM, b_i, b_k	Mkt, RMW, CMA
$\mathbb{M}_{90}^{(i,k)}$	SMB, RMW, CMA, MOM	Mkt, HML, b_i, b_k
$\mathbb{M}_{91}^{(i,k)}$	SMB, RMW, CMA, b_i, b_k	Mkt, HML, MOM
$\mathbb{M}_{92}^{(i,k)}$	SMB, RMW, MOM, b_i, b_k	Mkt, HML, CMA
$\mathbb{M}_{93}^{(i,k)}$	SMB, CMA, MOM, b_i, b_k	Mkt, HML, RMW
$\mathbb{M}_{94}^{(i,k)}$	HML, RMW, CMA, MOM	Mkt, SMB, b_i, b_k
$\mathbb{M}_{95}^{(i,k)}$	HML, RMW, CMA, b_i, b_k	Mkt, SMB, MOM
$\mathbb{M}_{96}^{(i,k)}$	HML, RMW, MOM, b_i, b_k	Mkt, SMB, CMA
$\mathbb{M}_{97}^{(i,k)}$	HML, CMA, MOM, b_i, b_k	Mkt, SMB, RMW
$\mathbb{M}_{98}^{(i,k)}$	RMW, CMA, MOM, b_i, b_k	Mkt, SMB, HML
$\mathbb{M}_{99}^{(i,k)}$	Mkt, SMB, HML, RMW, CMA	MOM, b_i, b_k
$\mathbb{M}_{100}^{(i,k)}$	Mkt, SMB, HML, RMW, MOM	CMA, b_i, b_k
$\mathbb{M}_{101}^{(i,k)}$	Mkt, SMB, HML, RMW, b_i, b_k	CMA, MOM
$\mathbb{M}_{102}^{(i,k)}$	Mkt, SMB, HML, CMA, MOM	RMW, b_i, b_k
$\mathbb{M}_{103}^{(i,k)}$	Mkt, SMB, HML, CMA, b_i, b_k	RMW, MOM
$\mathbb{M}_{104}^{(i,k)}$	Mkt, SMB, HML, MOM, b_i, b_k	RMW, CMA
$\mathbb{M}_{105}^{(i,k)}$	Mkt, SMB, RMW, CMA, MOM	HML, b_i, b_k
$\mathbb{M}_{106}^{(i,k)}$	Mkt, SMB, RMW, CMA, b_i, b_k	HML, MOM
$\mathbb{M}_{107}^{(i,k)}$	Mkt, SMB, RMW, MOM, b_i, b_k	HML, CMA
$\mathbb{M}_{108}^{(i,k)}$	Mkt, SMB, CMA, MOM, b_i, b_k	HML, RMW
$\mathbb{M}_{109}^{(i,k)}$	Mkt, HML, RMW, CMA, MOM	SMB, b_i, b_k
$\mathbb{M}_{110}^{(i,k)}$	Mkt, HML, RMW, CMA, b_i, b_k	SMB, MOM
$\mathbb{M}_{111}^{(i,k)}$	Mkt, HML, RMW, MOM, b_i, b_k	SMB, CMA
$\mathbb{M}_{112}^{(i,k)}$	Mkt, HML, CMA, MOM, b_i, b_k	SMB, RMW
$\mathbb{M}_{113}^{(i,k)}$	Mkt, RMW, CMA, MOM, b_i, b_k	SMB, HML
$\mathbb{M}_{114}^{(i,k)}$	SMB, HML, RMW, CMA, MOM	Mkt, b_i, b_k
$\mathbb{M}_{115}^{(i,k)}$	SMB, HML, RMW, CMA, b_i, b_k	Mkt, MOM
$\mathbb{M}_{116}^{(i,k)}$	SMB, HML, RMW, MOM, b_i, b_k	Mkt, CMA
$\mathbb{M}_{117}^{(i,k)}$	SMB, HML, CMA, MOM, b_i, b_k	Mkt, RMW
$\mathbb{M}_{118}^{(i,k)}$	SMB, RMW, CMA, MOM, b_i, b_k	Mkt, HML
$\mathbb{M}_{119}^{(i,k)}$	HML, RMW, CMA, MOM, b_i, b_k	Mkt, SMB
$\mathbb{M}_{120}^{(i,k)}$	Mkt, SMB, HML, RMW, CMA, MOM	b_i, b_k
$\mathbb{M}_{121}^{(i,k)}$	Mkt, SMB, HML, RMW, CMA, b_i, b_k	MOM
$\mathbb{M}_{122}^{(i,k)}$	Mkt, SMB, HML, RMW, MOM, b_i, b_k	CMA
$\mathbb{M}_{123}^{(i,k)}$	Mkt, SMB, HML, CMA, MOM, b_i, b_k	RMW
$\mathbb{M}_{124}^{(i,k)}$	Mkt, SMB, RMW, CMA, MOM, b_i, b_k	HML
$\mathbb{M}_{125}^{(i,k)}$	Mkt, HML, RMW, CMA, MOM, b_i, b_k	SMB
$\mathbb{M}_{126}^{(i,k)}$	SMB, HML, RMW, CMA, MOM, b_i, b_k	Mkt
$\mathbb{M}_{127}^{(i,k)}$	Mkt, SMB, HML, RMW, CMA, MOM, b_i, b_k	$\varnothing$

Table A.3 Candidate Splits for Factor Decomposition (Benchmark: FF6; three at a time). *Shading key:* gray (shaded) rows are models where $\{c_i, c_k, c_\ell\}$ is *jointly spanned* ($\{c_i, c_k, c_\ell\} \subset \boldsymbol{w}$); unshaded rows are models where it is *jointly unspanned* ($\{c_i, c_k, c_\ell\} \subset \boldsymbol{x}$).

Model	$\boldsymbol{x}$	$\boldsymbol{w}$
$\mathbb{M}_1^{(i,k,\ell)}$	Mkt	SMB, HML, RMW, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_2^{(i,k,\ell)}$	SMB	Mkt, HML, RMW, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_3^{(i,k,\ell)}$	HML	Mkt, SMB, RMW, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_4^{(i,k,\ell)}$	RMW	Mkt, SMB, HML, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_5^{(i,k,\ell)}$	CMA	Mkt, SMB, HML, RMW, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_6^{(i,k,\ell)}$	MOM	Mkt, SMB, HML, RMW, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_7^{(i,k,\ell)}$	c_i, c_k, c_ℓ	Mkt, SMB, HML, RMW, CMA, MOM
$\mathbb{M}_8^{(i,k,\ell)}$	Mkt, SMB	HML, RMW, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_9^{(i,k,\ell)}$	Mkt, HML	SMB, RMW, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{10}^{(i,k,\ell)}$	Mkt, RMW	SMB, HML, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{11}^{(i,k,\ell)}$	Mkt, CMA	SMB, HML, RMW, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{12}^{(i,k,\ell)}$	Mkt, MOM	SMB, HML, RMW, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{13}^{(i,k,\ell)}$	Mkt, c_i, c_k, c_ℓ	SMB, HML, RMW, CMA, MOM
$\mathbb{M}_{14}^{(i,k,\ell)}$	SMB, HML	Mkt, RMW, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{15}^{(i,k,\ell)}$	SMB, RMW	Mkt, HML, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{16}^{(i,k,\ell)}$	SMB, CMA	Mkt, HML, RMW, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{17}^{(i,k,\ell)}$	SMB, MOM	Mkt, HML, RMW, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{18}^{(i,k,\ell)}$	SMB, c_i, c_k, c_ℓ	Mkt, HML, RMW, CMA, MOM
$\mathbb{M}_{19}^{(i,k,\ell)}$	HML, RMW	Mkt, SMB, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{20}^{(i,k,\ell)}$	HML, CMA	Mkt, SMB, RMW, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{21}^{(i,k,\ell)}$	HML, MOM	Mkt, SMB, RMW, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{22}^{(i,k,\ell)}$	HML, c_i, c_k, c_ℓ	Mkt, SMB, RMW, CMA, MOM
$\mathbb{M}_{23}^{(i,k,\ell)}$	RMW, CMA	Mkt, SMB, HML, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{24}^{(i,k,\ell)}$	RMW, MOM	Mkt, SMB, HML, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{25}^{(i,k,\ell)}$	RMW, c_i, c_k, c_ℓ	Mkt, SMB, HML, CMA, MOM
$\mathbb{M}_{26}^{(i,k,\ell)}$	CMA, MOM	Mkt, SMB, HML, RMW, c_i, c_k, c_ℓ
$\mathbb{M}_{27}^{(i,k,\ell)}$	CMA, c_i, c_k, c_ℓ	Mkt, SMB, HML, RMW, MOM
$\mathbb{M}_{28}^{(i,k,\ell)}$	MOM, c_i, c_k, c_ℓ	Mkt, SMB, HML, RMW, CMA
$\mathbb{M}_{29}^{(i,k,\ell)}$	Mkt, SMB, HML	RMW, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{30}^{(i,k,\ell)}$	Mkt, SMB, RMW	HML, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{31}^{(i,k,\ell)}$	Mkt, SMB, CMA	HML, RMW, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{32}^{(i,k,\ell)}$	Mkt, SMB, MOM	HML, RMW, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{33}^{(i,k,\ell)}$	Mkt, SMB, c_i, c_k, c_ℓ	HML, RMW, CMA, MOM
$\mathbb{M}_{34}^{(i,k,\ell)}$	Mkt, HML, RMW	SMB, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{35}^{(i,k,\ell)}$	Mkt, HML, CMA	SMB, RMW, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{36}^{(i,k,\ell)}$	Mkt, HML, MOM	SMB, RMW, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{37}^{(i,k,\ell)}$	Mkt, HML, c_i, c_k, c_ℓ	SMB, RMW, CMA, MOM
$\mathbb{M}_{38}^{(i,k,\ell)}$	Mkt, RMW, CMA	SMB, HML, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{39}^{(i,k,\ell)}$	Mkt, RMW, MOM	SMB, HML, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{40}^{(i,k,\ell)}$	Mkt, RMW, c_i, c_k, c_ℓ	SMB, HML, CMA, MOM
$\mathbb{M}_{41}^{(i,k,\ell)}$	Mkt, CMA, MOM	SMB, HML, RMW, c_i, c_k, c_ℓ
$\mathbb{M}_{42}^{(i,k,\ell)}$	Mkt, CMA, c_i, c_k, c_ℓ	SMB, HML, RMW, MOM

continued on next page

Table A.3 (continued)

Model	$\boldsymbol{x}$	$\boldsymbol{w}$
$\mathbb{M}_{43}^{(i,k,\ell)}$	Mkt, MOM, c_i, c_k, c_ℓ	SMB, HML, RMW, CMA
$\mathbb{M}_{44}^{(i,k,\ell)}$	SMB, HML, RMW	Mkt, CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{45}^{(i,k,\ell)}$	SMB, HML, CMA	Mkt, RMW, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{46}^{(i,k,\ell)}$	SMB, HML, MOM	Mkt, RMW, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{47}^{(i,k,\ell)}$	SMB, HML, c_i, c_k, c_ℓ	Mkt, RMW, CMA, MOM
$\mathbb{M}_{48}^{(i,k,\ell)}$	SMB, RMW, CMA	Mkt, HML, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{49}^{(i,k,\ell)}$	SMB, RMW, MOM	Mkt, HML, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{50}^{(i,k,\ell)}$	SMB, RMW, c_i, c_k, c_ℓ	Mkt, HML, CMA, MOM
$\mathbb{M}_{51}^{(i,k,\ell)}$	SMB, CMA, MOM	Mkt, HML, RMW, c_i, c_k, c_ℓ
$\mathbb{M}_{52}^{(i,k,\ell)}$	SMB, CMA, c_i, c_k, c_ℓ	Mkt, HML, RMW, MOM
$\mathbb{M}_{53}^{(i,k,\ell)}$	SMB, MOM, c_i, c_k, c_ℓ	Mkt, HML, RMW, CMA
$\mathbb{M}_{54}^{(i,k,\ell)}$	HML, RMW, CMA	Mkt, SMB, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{55}^{(i,k,\ell)}$	HML, RMW, MOM	Mkt, SMB, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{56}^{(i,k,\ell)}$	HML, RMW, c_i, c_k, c_ℓ	Mkt, SMB, CMA, MOM
$\mathbb{M}_{57}^{(i,k,\ell)}$	HML, CMA, MOM	Mkt, SMB, RMW, c_i, c_k, c_ℓ
$\mathbb{M}_{58}^{(i,k,\ell)}$	HML, CMA, c_i, c_k, c_ℓ	Mkt, SMB, RMW, MOM
$\mathbb{M}_{59}^{(i,k,\ell)}$	HML, MOM, c_i, c_k, c_ℓ	Mkt, SMB, RMW, CMA
$\mathbb{M}_{60}^{(i,k,\ell)}$	RMW, CMA, MOM	Mkt, SMB, HML, c_i, c_k, c_ℓ
$\mathbb{M}_{61}^{(i,k,\ell)}$	RMW, CMA, c_i, c_k, c_ℓ	Mkt, SMB, HML, MOM
$\mathbb{M}_{62}^{(i,k,\ell)}$	RMW, MOM, c_i, c_k, c_ℓ	Mkt, SMB, HML, CMA
$\mathbb{M}_{63}^{(i,k,\ell)}$	CMA, MOM, c_i, c_k, c_ℓ	Mkt, SMB, HML, RMW
$\mathbb{M}_{64}^{(i,k,\ell)}$	Mkt, SMB, HML, RMW	CMA, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{65}^{(i,k,\ell)}$	Mkt, SMB, HML, CMA	RMW, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{66}^{(i,k,\ell)}$	Mkt, SMB, HML, MOM	RMW, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{67}^{(i,k,\ell)}$	Mkt, SMB, HML, c_i, c_k, c_ℓ	RMW, CMA, MOM
$\mathbb{M}_{68}^{(i,k,\ell)}$	Mkt, SMB, RMW, CMA	HML, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{69}^{(i,k,\ell)}$	Mkt, SMB, RMW, MOM	HML, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{70}^{(i,k,\ell)}$	Mkt, SMB, RMW, c_i, c_k, c_ℓ	HML, CMA, MOM
$\mathbb{M}_{71}^{(i,k,\ell)}$	Mkt, SMB, CMA, MOM	HML, RMW, c_i, c_k, c_ℓ
$\mathbb{M}_{72}^{(i,k,\ell)}$	Mkt, SMB, CMA, c_i, c_k, c_ℓ	HML, RMW, MOM
$\mathbb{M}_{73}^{(i,k,\ell)}$	Mkt, SMB, MOM, c_i, c_k, c_ℓ	HML, RMW, CMA
$\mathbb{M}_{74}^{(i,k,\ell)}$	Mkt, HML, RMW, CMA	SMB, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{75}^{(i,k,\ell)}$	Mkt, HML, RMW, MOM	SMB, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{76}^{(i,k,\ell)}$	Mkt, HML, RMW, c_i, c_k, c_ℓ	SMB, CMA, MOM
$\mathbb{M}_{77}^{(i,k,\ell)}$	Mkt, HML, CMA, MOM	SMB, RMW, c_i, c_k, c_ℓ
$\mathbb{M}_{78}^{(i,k,\ell)}$	Mkt, HML, CMA, c_i, c_k, c_ℓ	SMB, RMW, MOM
$\mathbb{M}_{79}^{(i,k,\ell)}$	Mkt, HML, MOM, c_i, c_k, c_ℓ	SMB, RMW, CMA
$\mathbb{M}_{80}^{(i,k,\ell)}$	Mkt, RMW, CMA, MOM	SMB, HML, c_i, c_k, c_ℓ
$\mathbb{M}_{81}^{(i,k,\ell)}$	Mkt, RMW, CMA, c_i, c_k, c_ℓ	SMB, HML, MOM
$\mathbb{M}_{82}^{(i,k,\ell)}$	Mkt, RMW, MOM, c_i, c_k, c_ℓ	SMB, HML, CMA
$\mathbb{M}_{83}^{(i,k,\ell)}$	Mkt, CMA, MOM, c_i, c_k, c_ℓ	SMB, HML, RMW
$\mathbb{M}_{84}^{(i,k,\ell)}$	SMB, HML, RMW, CMA	Mkt, MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{85}^{(i,k,\ell)}$	SMB, HML, RMW, MOM	Mkt, CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{86}^{(i,k,\ell)}$	SMB, HML, RMW, c_i, c_k, c_ℓ	Mkt, CMA, MOM
$\mathbb{M}_{87}^{(i,k,\ell)}$	SMB, HML, CMA, MOM	Mkt, RMW, c_i, c_k, c_ℓ

continued on next page

Table A.3 (continued)

Model	$\boldsymbol{x}$	$\boldsymbol{w}$
$\mathbb{M}_{88}^{(i,k,\ell)}$	SMB, HML, CMA, c_i, c_k, c_ℓ	Mkt, RMW, MOM
$\mathbb{M}_{89}^{(i,k,\ell)}$	SMB, HML, MOM, c_i, c_k, c_ℓ	Mkt, RMW, CMA
$\mathbb{M}_{90}^{(i,k,\ell)}$	SMB, RMW, CMA, MOM	Mkt, HML, c_i, c_k, c_ℓ
$\mathbb{M}_{91}^{(i,k,\ell)}$	SMB, RMW, CMA, c_i, c_k, c_ℓ	Mkt, HML, MOM
$\mathbb{M}_{92}^{(i,k,\ell)}$	SMB, RMW, MOM, c_i, c_k, c_ℓ	Mkt, HML, CMA
$\mathbb{M}_{93}^{(i,k,\ell)}$	SMB, CMA, MOM, c_i, c_k, c_ℓ	Mkt, HML, RMW
$\mathbb{M}_{94}^{(i,k,\ell)}$	HML, RMW, CMA, MOM	Mkt, SMB, c_i, c_k, c_ℓ
$\mathbb{M}_{95}^{(i,k,\ell)}$	HML, RMW, CMA, c_i, c_k, c_ℓ	Mkt, SMB, MOM
$\mathbb{M}_{96}^{(i,k,\ell)}$	HML, RMW, MOM, c_i, c_k, c_ℓ	Mkt, SMB, CMA
$\mathbb{M}_{97}^{(i,k,\ell)}$	HML, CMA, MOM, c_i, c_k, c_ℓ	Mkt, SMB, RMW
$\mathbb{M}_{98}^{(i,k,\ell)}$	RMW, CMA, MOM, c_i, c_k, c_ℓ	Mkt, SMB, HML
$\mathbb{M}_{99}^{(i,k,\ell)}$	Mkt, SMB, HML, RMW, CMA	MOM, c_i, c_k, c_ℓ
$\mathbb{M}_{100}^{(i,k,\ell)}$	Mkt, SMB, HML, RMW, MOM	CMA, c_i, c_k, c_ℓ
$\mathbb{M}_{101}^{(i,k,\ell)}$	Mkt, SMB, HML, RMW, c_i, c_k, c_ℓ	CMA, MOM
$\mathbb{M}_{102}^{(i,k,\ell)}$	Mkt, SMB, HML, CMA, MOM	RMW, c_i, c_k, c_ℓ
$\mathbb{M}_{103}^{(i,k,\ell)}$	Mkt, SMB, HML, CMA, c_i, c_k, c_ℓ	RMW, MOM
$\mathbb{M}_{104}^{(i,k,\ell)}$	Mkt, SMB, HML, MOM, c_i, c_k, c_ℓ	RMW, CMA
$\mathbb{M}_{105}^{(i,k,\ell)}$	Mkt, SMB, RMW, CMA, MOM	HML, c_i, c_k, c_ℓ
$\mathbb{M}_{106}^{(i,k,\ell)}$	Mkt, SMB, RMW, CMA, c_i, c_k, c_ℓ	HML, MOM
$\mathbb{M}_{107}^{(i,k,\ell)}$	Mkt, SMB, RMW, MOM, c_i, c_k, c_ℓ	HML, CMA
$\mathbb{M}_{108}^{(i,k,\ell)}$	Mkt, SMB, CMA, MOM, c_i, c_k, c_ℓ	HML, RMW
$\mathbb{M}_{109}^{(i,k,\ell)}$	Mkt, HML, RMW, CMA, MOM	SMB, c_i, c_k, c_ℓ
$\mathbb{M}_{110}^{(i,k,\ell)}$	Mkt, HML, RMW, CMA, c_i, c_k, c_ℓ	SMB, MOM
$\mathbb{M}_{111}^{(i,k,\ell)}$	Mkt, HML, RMW, MOM, c_i, c_k, c_ℓ	SMB, CMA
$\mathbb{M}_{112}^{(i,k,\ell)}$	Mkt, HML, CMA, MOM, c_i, c_k, c_ℓ	SMB, RMW
$\mathbb{M}_{113}^{(i,k,\ell)}$	Mkt, RMW, CMA, MOM, c_i, c_k, c_ℓ	SMB, HML
$\mathbb{M}_{114}^{(i,k,\ell)}$	SMB, HML, RMW, CMA, MOM	Mkt, c_i, c_k, c_ℓ
$\mathbb{M}_{115}^{(i,k,\ell)}$	SMB, HML, RMW, CMA, c_i, c_k, c_ℓ	Mkt, MOM
$\mathbb{M}_{116}^{(i,k,\ell)}$	SMB, HML, RMW, MOM, c_i, c_k, c_ℓ	Mkt, CMA
$\mathbb{M}_{117}^{(i,k,\ell)}$	SMB, HML, CMA, MOM, c_i, c_k, c_ℓ	Mkt, RMW
$\mathbb{M}_{118}^{(i,k,\ell)}$	SMB, RMW, CMA, MOM, c_i, c_k, c_ℓ	Mkt, HML
$\mathbb{M}_{119}^{(i,k,\ell)}$	HML, RMW, CMA, MOM, c_i, c_k, c_ℓ	Mkt, SMB
$\mathbb{M}_{120}^{(i,k,\ell)}$	Mkt, SMB, HML, RMW, CMA, MOM	c_i, c_k, c_ℓ
$\mathbb{M}_{121}^{(i,k,\ell)}$	Mkt, SMB, HML, RMW, CMA, c_i, c_k, c_ℓ	MOM
$\mathbb{M}_{122}^{(i,k,\ell)}$	Mkt, SMB, HML, RMW, MOM, c_i, c_k, c_ℓ	CMA
$\mathbb{M}_{123}^{(i,k,\ell)}$	Mkt, SMB, HML, CMA, MOM, c_i, c_k, c_ℓ	RMW
$\mathbb{M}_{124}^{(i,k,\ell)}$	Mkt, SMB, RMW, CMA, MOM, c_i, c_k, c_ℓ	HML
$\mathbb{M}_{125}^{(i,k,\ell)}$	Mkt, HML, RMW, CMA, MOM, c_i, c_k, c_ℓ	SMB
$\mathbb{M}_{126}^{(i,k,\ell)}$	SMB, HML, RMW, CMA, MOM, c_i, c_k, c_ℓ	Mkt
$\mathbb{M}_{127}^{(i,k,\ell)}$	Mkt, SMB, HML, RMW, CMA, MOM, c_i, c_k, c_ℓ	$\varnothing$

Table A.4 Equity Characteristics

Note: This table lists the description of 20 characteristics used in the empirical study.

No.	Characteristics	Description	Category
1	ABR	Abnormal returns around earnings announcement	Momentum
2	ACC	Operating accruals	Investment
3	ADM	Advertising expense-to-market	Intangibles
4	AGR	Asset growth	Investment
5	BASPREAD	Bid-ask spread (3 months)	Frictions
6	BETA	Beta (3 months)	Frictions
7	BM	Book-to-market equity	Value-versus-growth
8	CFP	Cashflow-to-price	Value-versus-growth
9	EP	Earnings-to-price	Value-versus-growth
10	ME	Market equity	Frictions
11	MOM1M	Previous month return	Momentum
12	MOM12M	Cumulative returns in the past (2-12) months	Momentum
13	NI	Net equity issue	Investment
14	OP	Operating profitability	Profitability
15	RDM	R&D-to-market	Intangibles
16	ROE	Return on equity	Profitability
17	SEAS1A	1-Year Seasonality	Intangibles
18	SP	Sales-to-price	Value-versus-growth
19	SUE	Standardized unexpected quarterly earnings	Momentum
20	SVAR	Return variance (3 months)	Frictions

Figure A.1 Counts of Fake Anomalies identified under different q thresholds (in $q = 0.2$ case)

This figure plots, for Steps 1-6, the number of fake anomalies identified when applying EFDR thresholds of $q = 0.10$, $q = 0.15$, and $q = 0.20$, using the EFDR distributions obtained under the baseline configuration with $q = 0.20$.

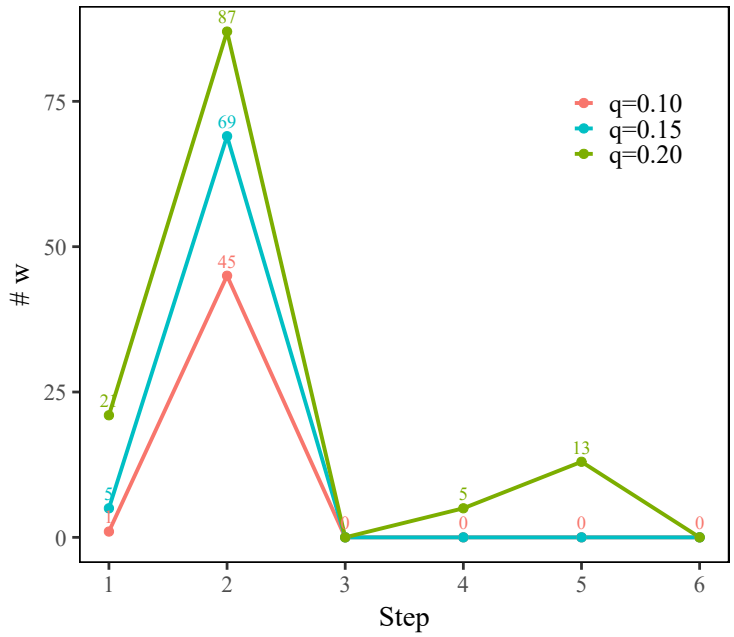


Figure A.2 Counts of Fake Anomalies and Remaining Anomalies. (Benchmark: `FF6`, $q = 0.1$).

Note: This figure reports the counts of anomalies classified as $\boldsymbol{w}$ and $\boldsymbol{x}$ by our procedure, using `FF6` as the benchmark. Green bars denote remaining anomalies ($\boldsymbol{x}$), and red bars denote fake anomalies ($\boldsymbol{w}$). The threshold q is 0.1.