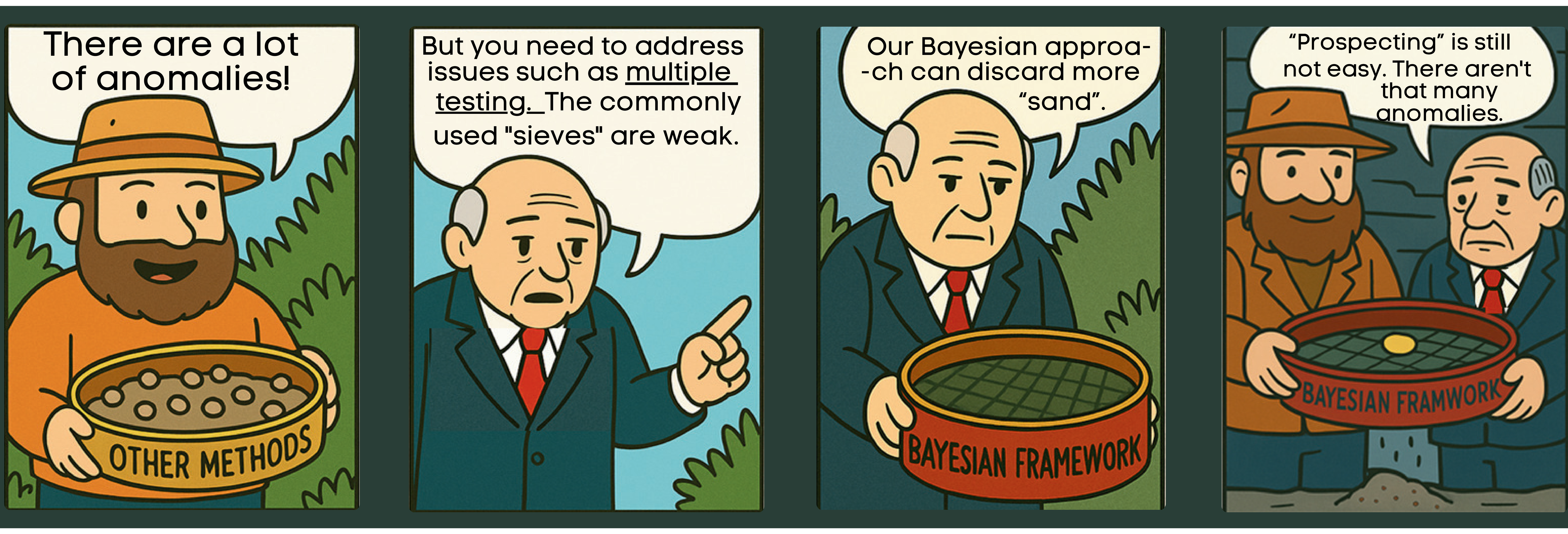
Factors or Fake? A New Look at Anomalies and the Replication Crisis

Siddhartha Chib — Washington University in St. Louis
Lingxiao Zhao — Peking University
Shuhua Xiao — **City University of Hong Kong**



"Gold" panning

Marginal Likelihood is all you need !

Methodology

Single Testing

$\mathcal{F}_i$: normal "risk factor", e.g, Mkt, FF3, FF6
$\boldsymbol{a}_i$: candidate "anomaly"
$\boldsymbol{f}_i = (\mathcal{F}_i, \boldsymbol{a}_i)$: set of factors, consist with Mkt/FF3/FF6 + 1 anomaly

$\boldsymbol{f}$ can be decomposed as risk factors $\boldsymbol{x}$ and non-risk factors $\boldsymbol{w}$ by:
$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{\lambda}_x + \boldsymbol{u}_{x,t} \\ \boldsymbol{w}_t = \Gamma \boldsymbol{x}_t + \boldsymbol{\varepsilon}_{w\cdot x,t} \end{cases}$$

$\boldsymbol{f} \to (\boldsymbol{x}, \boldsymbol{w})$

1. Enumerate all possible configurations $\mathcal{M}$
2. Compute marginal likelihood (ML) for each
3. "true model" : model with the highest ML $\widehat{M}$

If considering data segmentation, in each regime s:
$$\boldsymbol{x}_{j,s,t} = \boldsymbol{\lambda}_{x,j,s} + \boldsymbol{u}_{x,j,s,t}, \qquad \boldsymbol{w}_{j,s,t} = \boldsymbol{\Gamma}_{j,s}\boldsymbol{x}_{j,s,t} + \boldsymbol{\varepsilon}_{w,j,s,t},$$
optimal breakpoint: breakpoint with the highest ML

Anomaly: cannot be fully explained by existing risk factors

Once an anomaly is identified, it should be re-integrated into the stochastic discount factor (SDF). It should be the risk factor.

$$\begin{cases} \text{If } \boldsymbol{a} \in \boldsymbol{w}, \text{ then it is not an anomaly.} \\ \text{If } \boldsymbol{a} \in \boldsymbol{x}, \text{ then it is an anomaly.} \end{cases}$$

Prop1. Connection Between the Alpha Test and Bayesian Factor Classification

Alpha Test: $a_{it} = \alpha_i + \boldsymbol{b}_i' \mathcal{F}_t + \varepsilon_{it}$, (Model $\mathcal{M}_\alpha$)
Factor deco.: $\boldsymbol{f} = (\mathcal{F}, \boldsymbol{a})$

$\mathcal{M}_A \subset \mathcal{M}$, subset of models in which $\boldsymbol{a} \in \boldsymbol{x}$
$\mathcal{M}_A^c = \mathcal{M} \setminus \mathcal{M}_A$, subset of models in which $\boldsymbol{a} \in \boldsymbol{w}$

- If the null is rejected in model $\mathcal{M}_\alpha$ then $\widehat{M} \in \mathcal{M}_A$
- If the null is not rejected, then $\widehat{M} \in \mathcal{M}_A^c$

Rather than merely accepting or rejecting a null hypothesis, the marginal likelihood approach provides a posterior ranking over competing models, enabling us to evaluate the relative plausibility of each specification.

Sequential Multiple Testing

Local False Discovery Rate (lfdr)

lfdr for $\boldsymbol{a}_i$: the posterior prob. that it is not anomalous

$$\begin{cases} \text{For each } \boldsymbol{a}_i,\ J \text{ competing models:} & M_1^{(i)}, \dots, M_J^{(i)} \\ \text{Posterior model prob. for model } M_j^{(i)}: & \mathbb{P}(M_j^{(i)} \mid \text{data}) \\ \boldsymbol{a}_i \in \boldsymbol{w}, \text{ the indices of these models:} & \mathcal{N} \subseteq \{1, \dots, J\} \end{cases}$$

$$\text{lfdr}_i = \sum_{j \in \mathcal{N}} \mathbb{P}(M_j^{(i)} \mid \text{data}) = \mathbb{P}(\boldsymbol{a}_i \in \boldsymbol{w} \mid \text{data})$$

Adaptive Thresholding to Control Expected FDR (EFDR)

sort lfdr_i in ascending order:
$$\text{lfdr}_{(1)} \le \text{lfdr}_{(2)} \le \cdots \le \text{lfdr}_{(n)}$$

calculate the expected FDR:
$$\text{FDR}(k) = 1/k \sum_{\ell=1}^{k} \text{lfdr}_{(\ell)}$$

optimal threshold: $k^* = \max\{k : \text{FDR}(k) \le q\}$

$$\begin{cases} (n-k^*) \text{ assets} \sim \text{lfdr}_{(k^*+1)}, \dots, \text{lfdr}_{(n)}, \text{ which are priced} \\ k^* \text{ assets} \sim \text{lfdr}_{(1)}, \dots, \text{lfdr}_{(k^*)}, \text{ which are NOT priced} \end{cases}$$

Anomaly Pricing: Two (Three, Four, ...) at a Time

Anomalies left unpriced from step 1:
$$\boldsymbol{b}_1, \dots, \boldsymbol{b}_m, m = k^*$$
Augmented factor set
$$\boldsymbol{f}_{j,k} = \{\boldsymbol{b}_j, \boldsymbol{b}_k\} \cup \mathcal{F}$$
Average marginal inclusion prob. for each anomaly:
$$\bar{p}_{(k)}^* = \frac{1}{j^* - 1} \sum_{k \ne j} \Pr(\boldsymbol{b}_j \in \boldsymbol{w} \mid \text{data}, \boldsymbol{f}_{j,k}).$$
Sort $\bar{p}_{(j)}^*$; find the largest j^{**} such that $1/j^{**} \sum_{j=1}^{j^{**}} \bar{p}_{(j)} \le q$.

$$\begin{cases} b_{(j^{**}+1)}, \dots, b_{(m)} \sim \text{priced} \\ b_{(1)}, \dots, b_{(j^{**})} \sim \text{NOT priced} \end{cases}$$

If full set: $\boldsymbol{f} = (\mathcal{F}, \text{all } \boldsymbol{a}_i)$, computationally intensive!

- Start with: $\boldsymbol{f}_i = \{\boldsymbol{a}_i\} \cup \mathcal{F}$, filter out priced components
- Move to: $\boldsymbol{f}_{j,k} = \{\boldsymbol{b}_j, \boldsymbol{b}_k\} \cup \mathcal{F}$, filter out priced components
- ... until the set of remaining (non-)anomalies stabilizes

"Gold" panning: some noise (non-anomalies) is removed, but noise may still remain due to the unobservable nature of all true factors.

Algorithm Stepwise Local FDR Filtering Algorithm

Require: Dataset $\mathcal{D}$ of T observations; Anomaly set $\mathcal{A}$; Benchmark factor set $\mathcal{F}$; FDR threshold $q \in (0,1)$. Training proportion $\rho \in (0,1)$;
Ensure: Final set of anomalies deened *not priced*, $\widetilde{\mathcal{A}}$.

```
1 : T ← |D|
2 : T_train ← ⌊ρ · T⌋
3 : D_train ← first T_train rows of D
4 : Initialize Ã ← A
5 : for n ← 1 to |A| do
6 :     if |Ã| < n then break
7 :     end if
8 :     for each anomaly a_j ∈ Ã do
9 :         S ← Ã \ {a_j}
10 :        sumProb ← 0
11 :        for each combination G ⊆ S with |G| = n − 1 do
12 :            X ← F ∪ {a_j} ∪ G
13 :            Scan all possible models of X, calculate marginal
               likelihood for each to obtain posterior probability
               p_{j,G} = Pr(a_j unpriced | X)
14 :            sumProb ← sumProb + p_{j,G}
15 :        end for
16 :        lfdr^(a_j) ← sumProb / C(|S|, n − 1)
17 :    end for
18 :    Sort anomalies in Ã by increasing lfdr^, yielding
           ℓ_(1) ≤ ℓ_(2) ≤ ··· ≤ ℓ_(|Ã|)
19 :    Find the largest index k* such that
           (1/k*) Σ_{i=1}^{k*} ℓ_(i) ≤ q
20 :    Update Ã ← {a_j : lfdr^(a_j) ≤ ℓ_(k*)}.
21 : end for
22 : return Ã
```

Empirical Results

Data: Nov 1971 ~ Dec 2024, 153 U.S. factors from Jensen, Kelly, and Pedersen (2023, JF)

Table1: Proportion of Non-Anomalies Under Single Testing

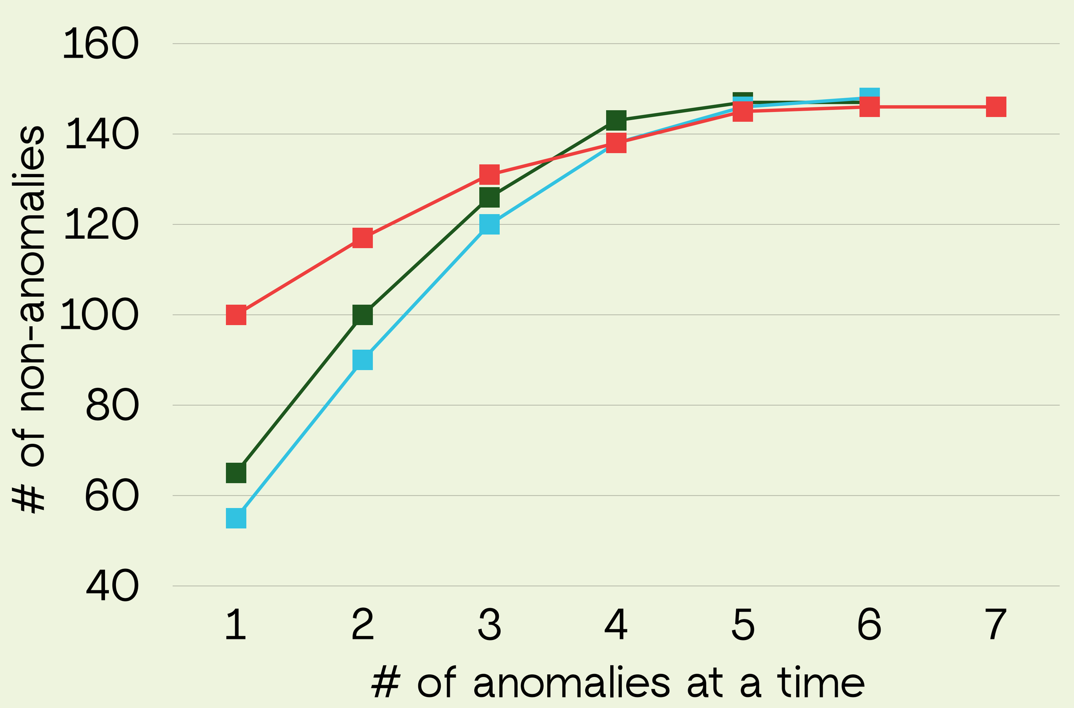
	No break	One break		Two breaks		
	Whole	Regime1	Regime2	Regime1	Regime2	Regime3
Mkt	40.5	45.8	55.6	49.0	45.1	65.4
FF3	33.3	39.2	62.7	45.8	51.6	64.7
FF6	62.1	58.2	56.9	58.2	52.3	66.7

Table2: Proportion of Non-Anomalies Under Multiple Testing

	No break	One break		Two breaks		
	Whole	Regime1	Regime2	Regime1	Regime2	Regime3
Mkt	42.5	46.4	68.0	54.3	69.3	77.8
FF3	36.0	39.9	68.6	45.8	73.2	75.2
FF6	65.4	73.2	71.9	73.9	85.0	85.6

One-at-a-time, q=0.1

Figure1: Number of Non-Anomalies Identified Under Sequential Multiple Testing and Different $\mathcal{F}_i$



Feel free to follow my homepage for the upcoming draft.

Highlights / Summary

- A Bayesian framework that incorporates both single and multiple testing to efficiently screen out candidate anomalies that are inevitably priced (not anomalies).
- In single testing, our approach parallels the traditional alpha test, which can also be implemented with regime identification.
- In multiple testing, we propose a Bayesian false discovery rate procedure. Result of the sequential process is closed to the full scan.
- Many "anomalies" arise from ignoring risk factor instability / model misspecification / multiple testing / power of testing methods.