Is There A Replication Crisis in Finance?

Theis Ingerslev Jensen
CBS

Bryan Kelly
Yale, AQR

Lasse Heje Pedersen
CBS, AQR

Finance Faces a Replication Crisis[†]

Challenges to the replicability of factor research take two basic forms:

1. **No internal validity.** Main results cannot be replicated using slightly different methodologies or data.

   E.g., Hou et al. (2020) state: *"Most anomalies fail to hold up to currently acceptable standards for empirical finance"*

2. **No external validity.** Results replicate in-sample, but are spurious and driven by "*p*-hacking." Sheer number of factors is too large to be believable. E.g., Cochrane (2011) asks for a consolidation of the "factor zoo," and Harvey and Liu (2016) state: *"most claimed research findings in financial economics are likely false."*

[†] And many other fields: Ioannidis (2005) "Why most published research findings are false" *PLoS Medicine*

What We Do: Theory-based Replication

Question: What fraction of factor research is replicable?

What We Do: Theory-based Replication

Question: What fraction of factor research is replicable?

Answer: 85%

What We Do: Theory-based Replication

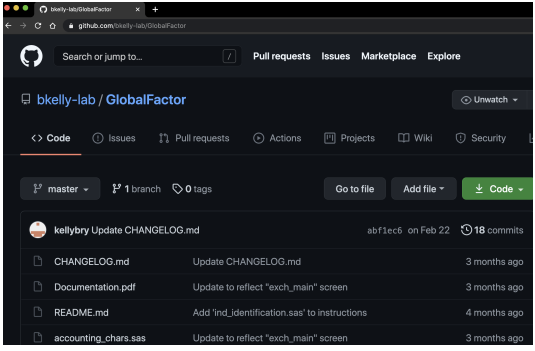
Question: What fraction of factor research is replicable?

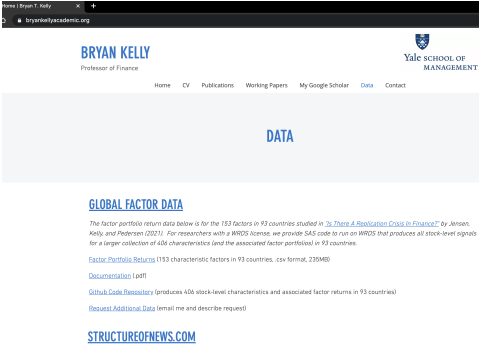
Answer: 85%

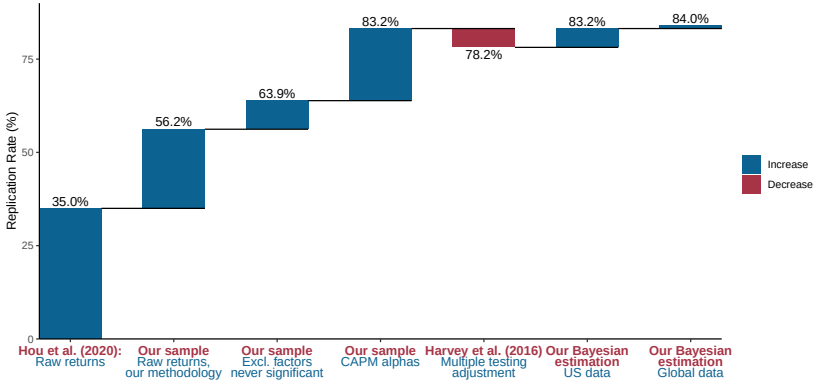
Based on

- Theory-based Bayesian approach
  - Economic theory
  - Model for logical learning about replication
  - Multiple testing correction
- Large new replicable data set
  - 153 factors across 93 countries, constructed in a simple consistent way
  - Code and data publicly available

Data and Code







Factor construction differences

- 1 month holding period (vs. 1, 6, and 12 month) (+4.0%)
- "Capped" value weights (+8.5%)

Sample differences

- Exclude factors that were never significant (+7.8%)
- Longer time series (+4.3%)
- Global data (+0.9%)

Method differences

- CAPM alpha (vs. raw returns) (+20.9%)
- Hierarchical model joint estimation (vs. independent tests)
- Bayesian prior (vs. frequentist multiple testing correction) (-0.9%)

Bayesian Model

A Single Factor

Bayesian prior is CAPM holds $f_t = \alpha + \beta r_t^m + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \quad \alpha \sim N(0, \tau^2)$

Denoting $\hat{\alpha} = \frac{1}{T} \sum_t (f_t - \beta r_t^m),$

Posterior normal with $E(\alpha|\hat{\alpha}) = \kappa\hat{\alpha}$ where $\kappa = \frac{1}{1+\frac{\sigma^2}{\tau^2 T}} \in (0, 1)$

Bayesian Model

A Single Factor

Bayesian prior is CAPM holds $f_t = \alpha + \beta r_t^m + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \quad \alpha \sim N(0, \tau^2)$

Denoting $\hat{\alpha} = \frac{1}{T} \sum_t (f_t - \beta r_t^m),$

Posterior normal with $E(\alpha|\hat{\alpha}) = \kappa\hat{\alpha}$ where $\kappa = \frac{1}{1+\frac{\sigma^2}{\tau^2 T}} \in (0, 1)$

- A positive, but lower, alpha sometimes interpreted as sign of replication failure
- But it is expected outcome from Bayesian perspective
- Decline in post-publication factor performance (McLean and Pontiff 2016) in line with posterior a Bayesian would have formed from published results

*Note: Paper considers an extended Bayesian model that allows for "alpha-hacking"

Bayesian Model

Hierarchical Alphas

- Factors are correlated and conceptually related to each other
- "Domestic" ($f_t = \alpha + \beta r_t^m + \varepsilon_t$) plus "global" ($f_t^g = \alpha + \beta^g r_t^g + \varepsilon_t^g$) evidence

Proposition (The Power of Shared Evidence)

The posterior alpha given domestic ($\hat{\alpha}$) and global ($\hat{\alpha}^g$) evidence is normal with

$$E(\alpha|\hat{\alpha}, \hat{\alpha}^g) = \kappa^g \left(\frac{1}{2}\hat{\alpha} + \frac{1}{2}\hat{\alpha}^g \right)$$

Less shrinkage and more conviction

$$\kappa^g = \frac{1}{1 + \frac{\sigma^2}{\tau^2 T}\frac{1+\rho}{2}} \in [\kappa, 1], \qquad Var(\alpha|\hat{\alpha}) \geq Var(\alpha|\hat{\alpha}, \hat{\alpha}^g)$$

Bayesian Model

Model

- $f_t = \alpha^i + \varepsilon_t^i, \qquad \alpha^i = \alpha^o + c^j + \omega^i, \qquad \alpha^o = 0,\ c^j \sim N(0, \tau_c^2), w^i \sim N(0, \tau_w^2)$
- Global analysis adds another tier to hierarchy

Estimation

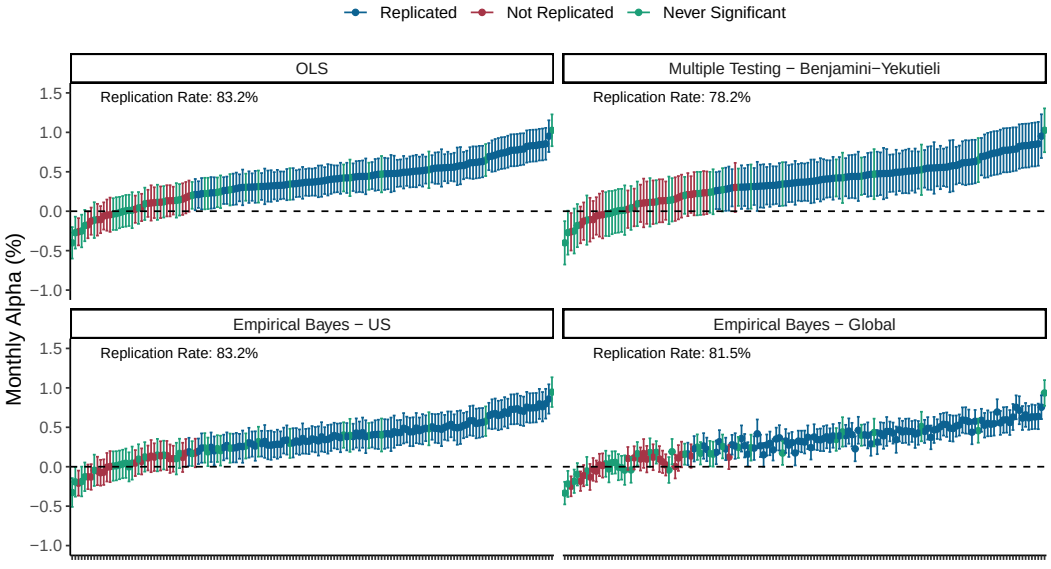
- Empirical Bayes
- Intuition: Realized dispersion in $\hat{\alpha}^i$'s can inform prior

Bayesian Multiple Testing

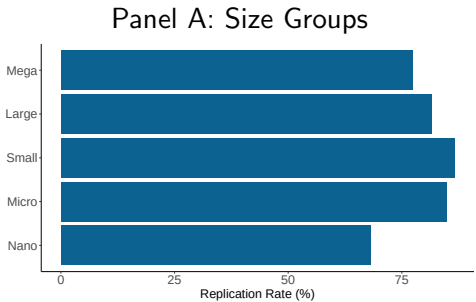
- Controls false discoveries, yet preserves power (c.f. frequentist corrections)
- From posterior, can make *any* inference calculation (posterior of null, FDR, FWER, ...)
- *"The problem of multiple comparisons can disappear entirely when viewed from a hierarchical Bayesian perspective." Gelman et al. (2012)*

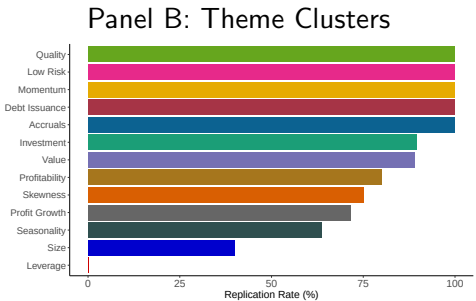
Empirical Results

Internal Validity

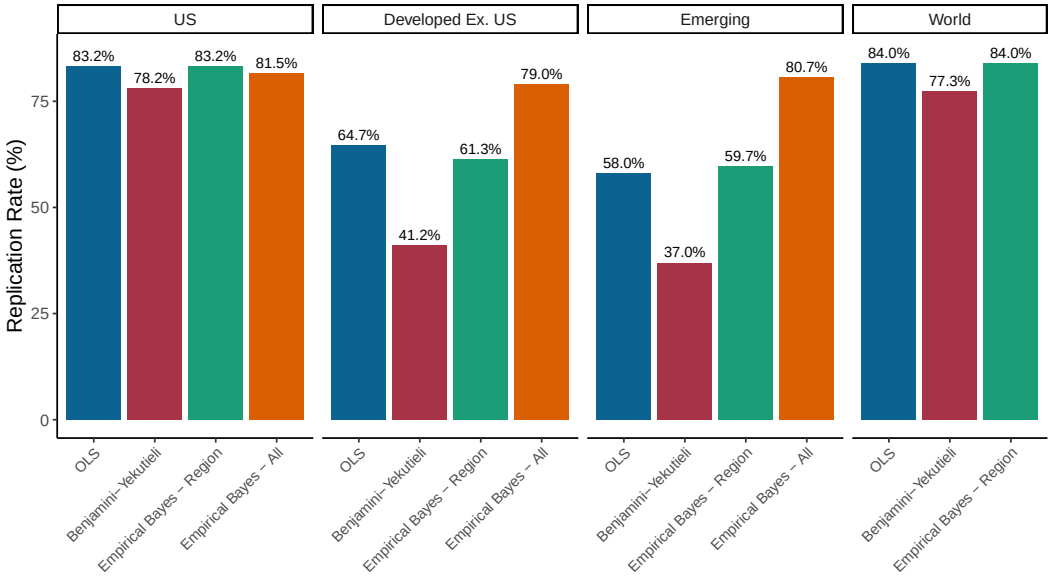


Internal Validity

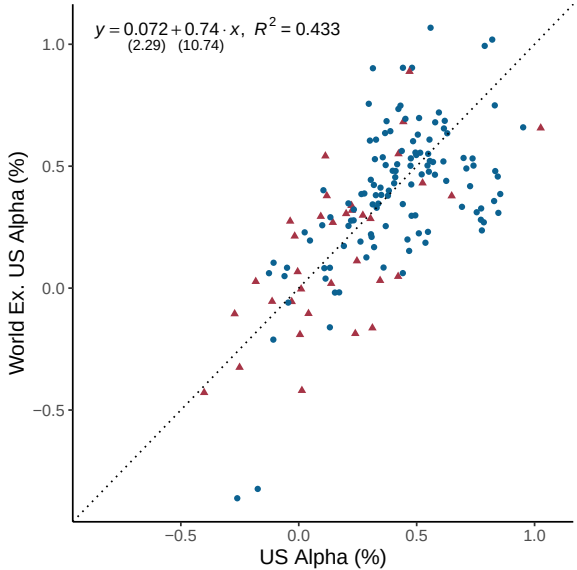




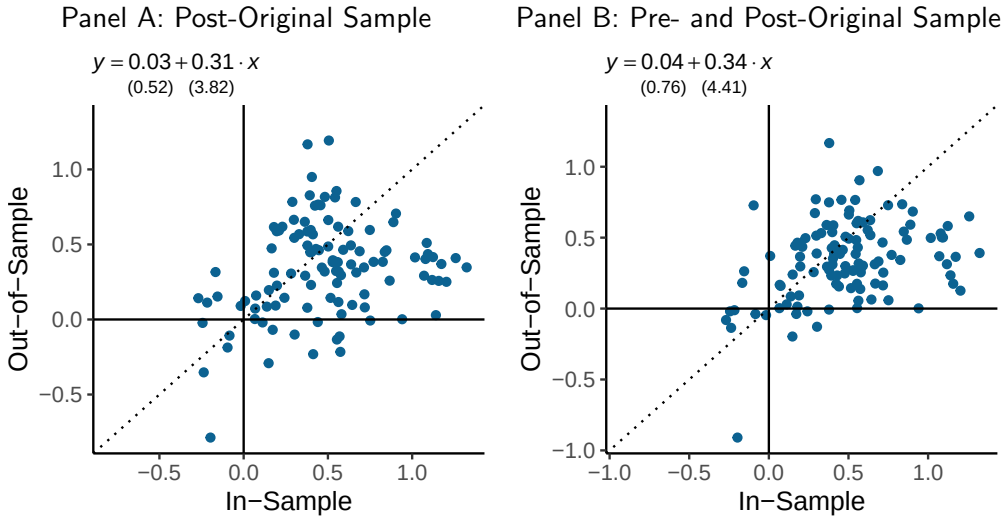
External Validity: Global



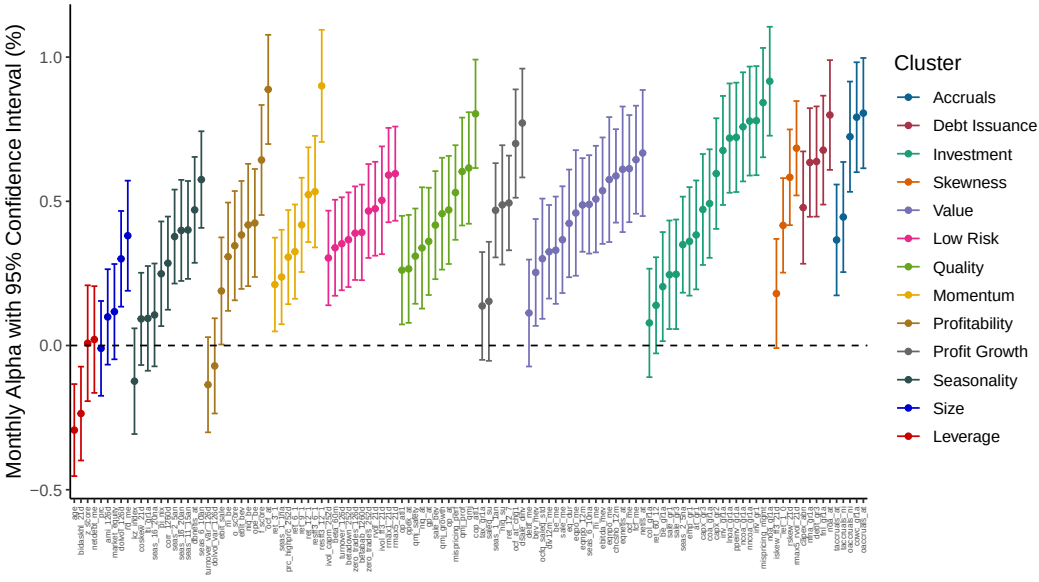
External Validity: Global



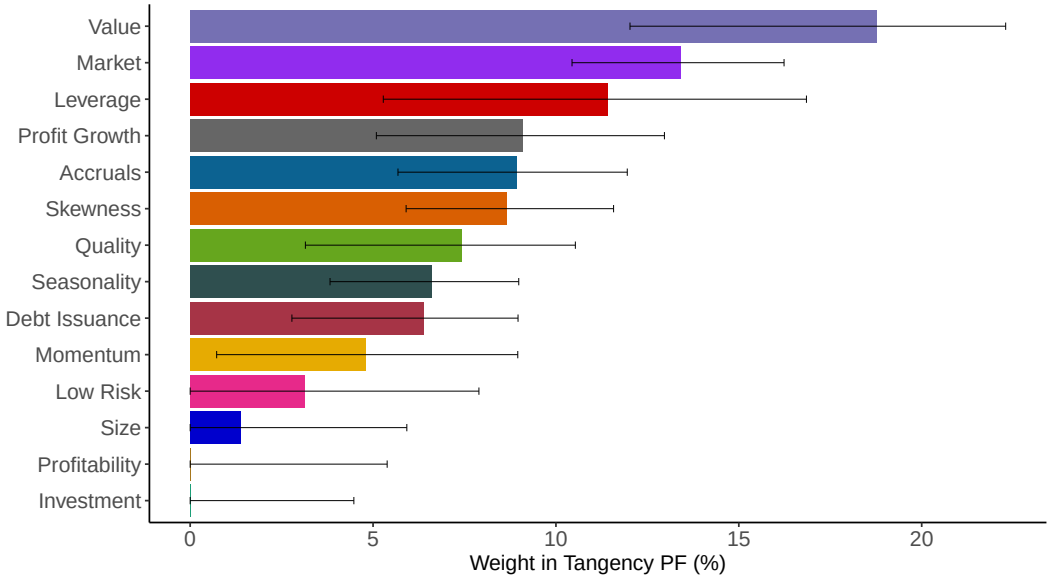
External Validity: Time Series



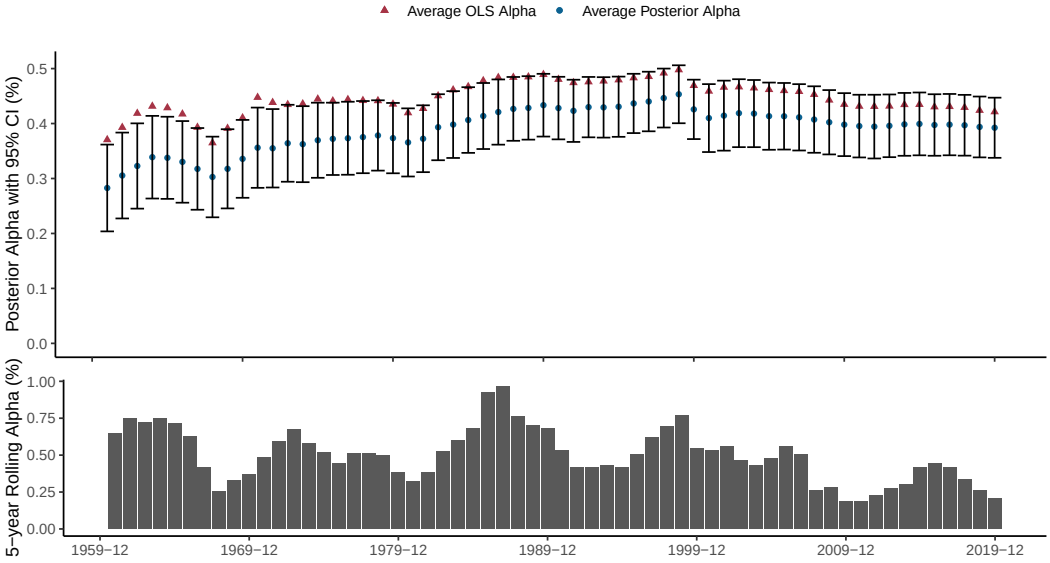
Economic Significance: Individual Factors



Economic Significance: Which Factors Matter *Jointly*?



Bayesian Posterior In Real Time



Conclusion: Finance Research Posterior

- Factor research exhibits high degree of internal and external validity
- 85% replication rate in global sample over long history
- Introduce hierarchical Bayesian model of alphas that
  - emphasizes the joint behavior of factors
  - more powerful multiple test adjustment than common frequentist methods
- Post-publication factor decay is closely in line with Bayesian posteriors based on publication evidence
  - Post-pub data largely confirms Bayesian's beliefs $\Rightarrow$ stable alpha posterior over time
- Our code, data, and documentation are available online
  - Updated regularly with the new data releases and bug fixes
  - `https://github.com/bkelly-lab/GlobalFactor`