

# 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?

**Answer:** 82.4%

Based on

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

# Replication Rate Comparison

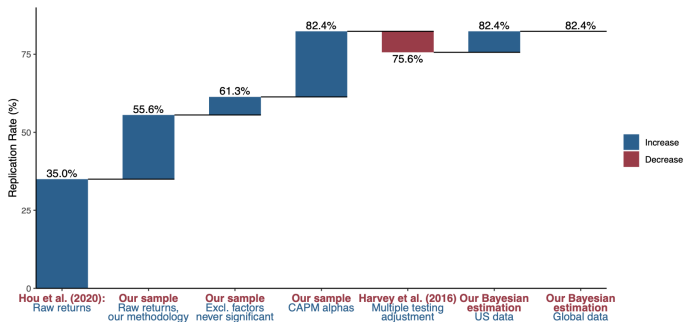


Figure 1: Replication Rates Versus the Literature

# Summary of Replication Rate Differences

## 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

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

Posterior normal with

$$E(\alpha|\hat{\alpha}) = \kappa \hat{\alpha}, \quad \text{Var}(\alpha|\hat{\alpha}) = \frac{1}{\frac{1}{\sigma^2/T} + \frac{1}{\tau^2}}$$

where

$$\hat{\alpha} = \frac{1}{T} \sum_t (f_t - \beta r_t^m) \quad \text{and} \quad \kappa = \frac{1}{1 + \frac{\sigma^2}{\tau^2 T}} \in (0, 1)$$

- ▶ 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

# Bayesian Model

## Related Factors

- ▶ “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

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

$$\kappa^g = \frac{1}{1 + \frac{\sigma^2}{\tau^2 T} \frac{1+\rho}{2}} \in [\kappa, 1]$$

Less shrinkage More conviction

$$\text{Var}(\alpha) \geq \text{Var}(\alpha|\hat{\alpha}, \hat{\alpha}^g)$$

# Bayesian Model

## Model

- ▶  $f_t = \alpha' + \varepsilon'_t$ ,  $\alpha' = \alpha^0 + c' + \omega'$ ,  $\alpha^0 = 0$ ,  $c' \sim N(0, \tau_c^2)$ ,  
 $\omega' \sim N(0, \tau_\omega^2)$
- ▶ Global analysis adds another tier to hierarchy

## Estimation

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

## Bayesian Multiple Testing

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



# Internal Validity

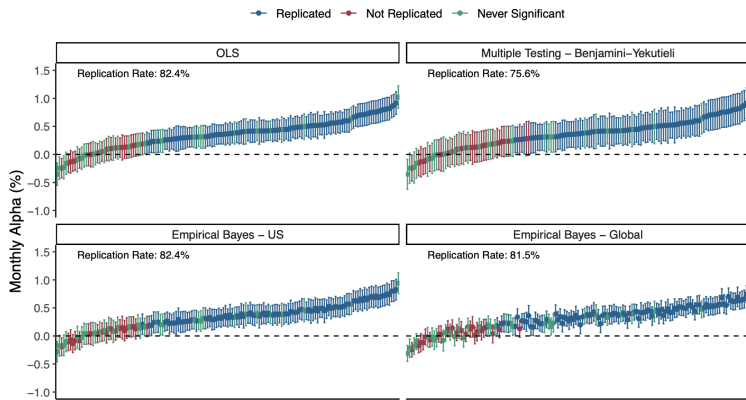
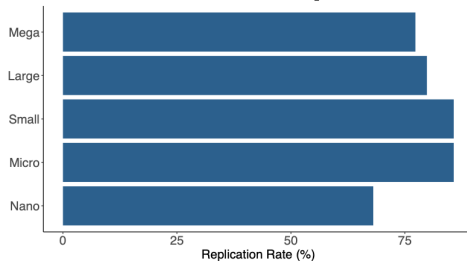


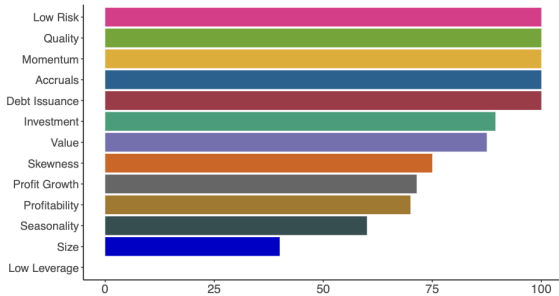
Figure 4: Alpha Distributions for US Factors

# Internal Validity

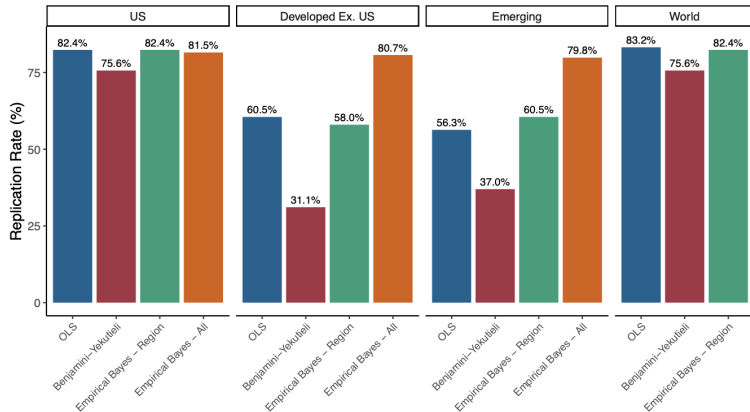
Panel A: Size Groups



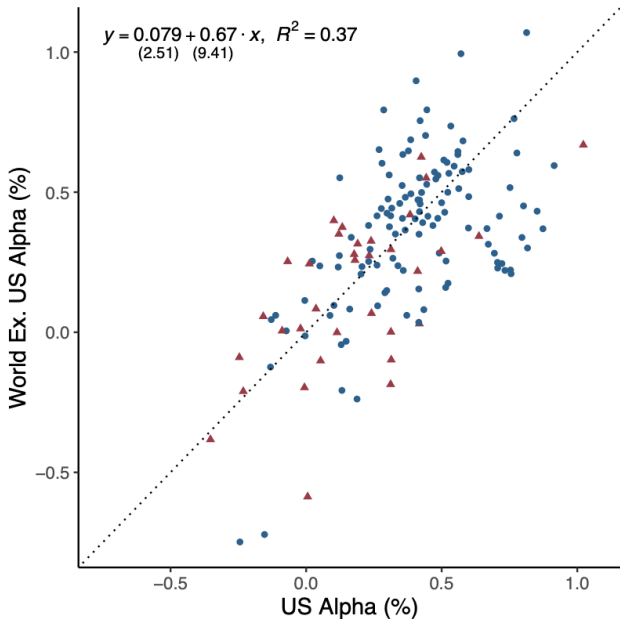
Panel B: Theme Clusters



# External Validity: Global

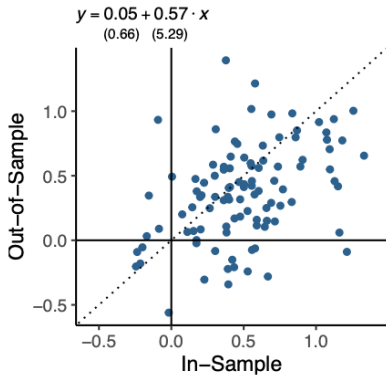


# External Validity: Golbal

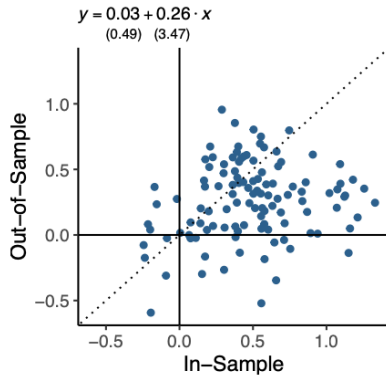


# External Validity: Time Series

Panel A: Pre-original Sample



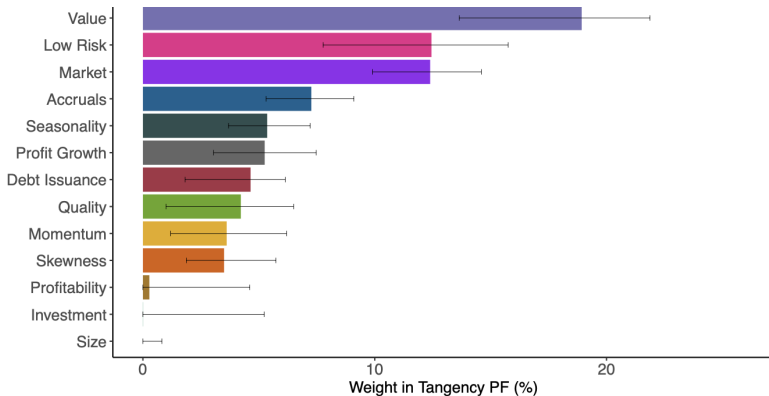
Panel B: Post-original Sample



# Economic Significance: Individual Factors

ES: .png

# Economic Significance: Which Factors Matter Jointly?



# Conclusion: Finance Research Posterior

- ▶ Factor research exhibits high degree of internal and external validity
- ▶ 82.4% 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>



# Comparison with Hou et al 2020

Hou et al is better.

- ▶ They keep sample simple. This paper use too much tricks
- ▶ They keep model simple. This paper is some kind of model-driven research. It's more like a paper of stats or even computer science.