# Testing Weak Factors in Asset Pricing

Soohun Kim  Valentina Raponi  Paolo Zaffaroni

NBER

11 July 2025

KAIST

IESE

Imperial College London and Sapienza Rome

# Outline

Overview

Economy

- Conditional Asset Pricing Set-Up
- Benchmark Case: No Strong Factor
- Observed Strong Factors
- Unobserved Strong Factors

Simulation

Empirical Application

Conclusion

# Empirical Asset Pricing

- One of the most famous equations in AP is

$$\mu(\text{rewards}) = B(\text{risk}) \times \gamma(\text{rewards per unit risk})$$

# Empirical Asset Pricing

- One of the most famous equations in AP is

$$\mu(\text{rewards}) = B(\text{risk}) \times \gamma(\text{rewards per unit risk})$$

- Seemingly benign but captivating
  - Standard empirical approach is Two-Pass CSR method
  - Once you decide to take it seriously, lots of complexity arises in empirical application

# Empirical Asset Pricing

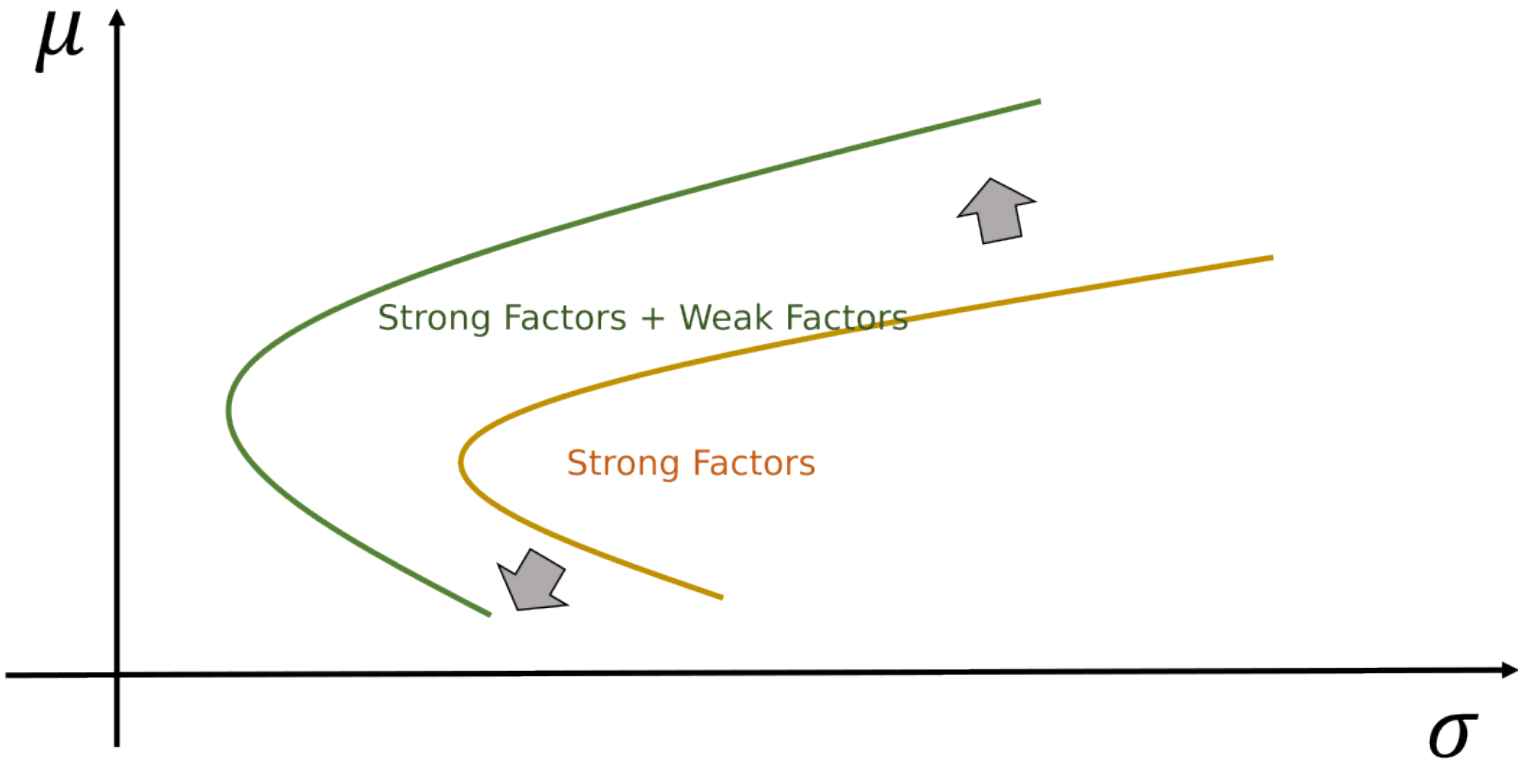
- One of the most famous equations in AP is

$$\mu(\text{rewards}) = B(\text{risk}) \times \gamma(\text{rewards per unit risk})$$

- Seemingly benign but captivating
  - Standard empirical approach is Two-Pass CSR method
  - Once you decide to take it seriously, lots of complexity arises in empirical application
- This paper considers the issue of weak factors
  - When factor loadings are close to zero for most assets
  - PCA will confound them with noise

# Weak Factors and Investment

- APT is fine with the following Mean-Variance analysis:




- Hence, as an investor, she will have a strong incentive to search for the weak factors!

# Weak Factors and Asset Pricing Test

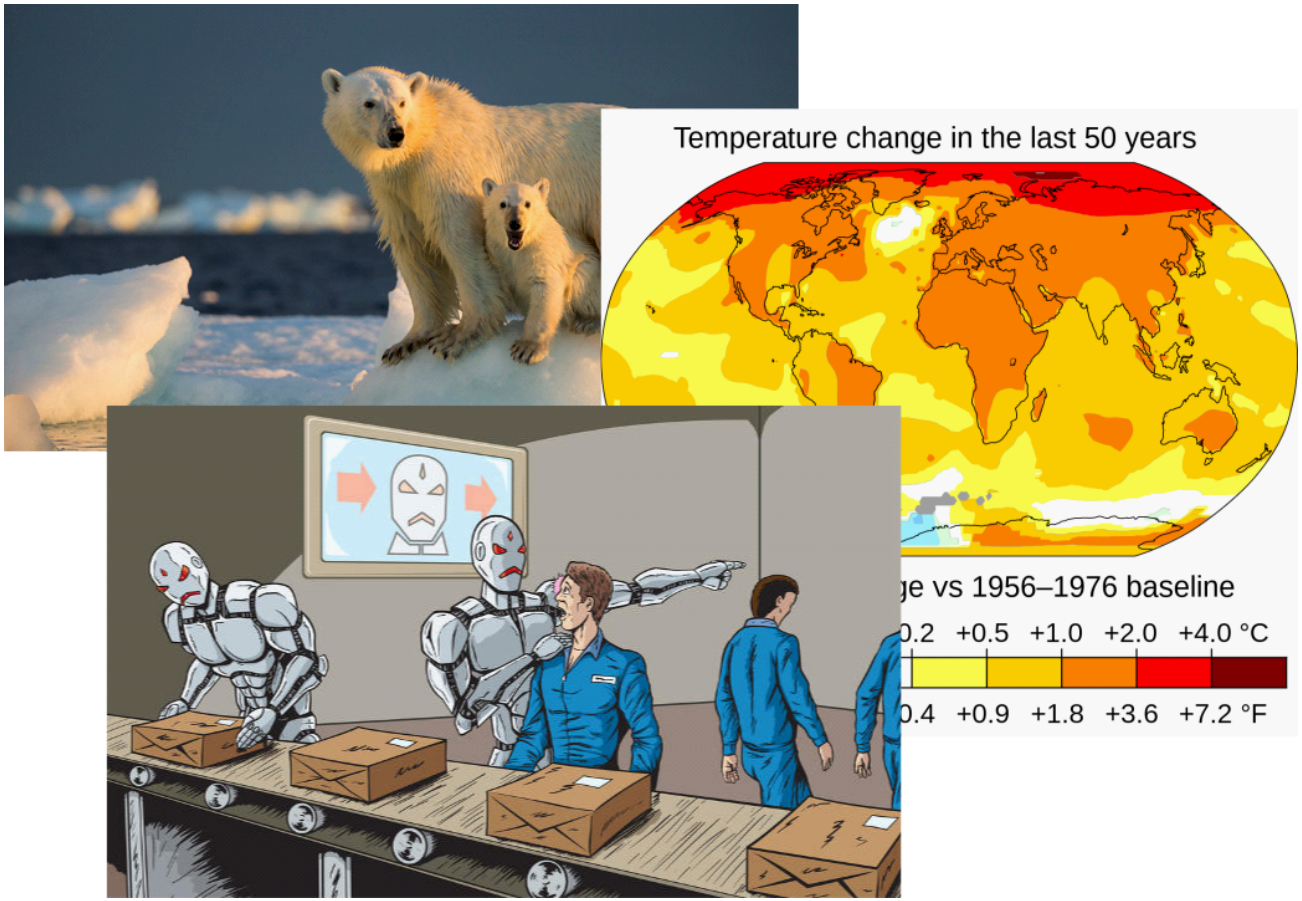
- When some factors are weak, lots of distortion may happen
  - weak factors without premium may appear to be important
  - strong factors with significant premium may appear insignificant

# Weak Factors and Asset Pricing Test

- When some factors are weak, lots of distortion may happen
  - weak factors without premium may appear to be important
  - strong factors with significant premium may appear insignificant
- Especially, when the literature proposes hundreds of factors, we need some criteria

# Furthermore, rapidly changing economic landscape

- We need to discern which factors are strong weak
  - in a rapidly changing economic environment
- For example, paradigm shifts such as climate changes or job destruction due to AI beg for a short-$T$ method



# Origins of Weak Factors

Weak factors may emerge...

- ...when constructing factors based on anomalies or when considering macroeconomic factors, as they are typically exposed to a smaller subset of the test assets under examination.
- ...under market incompleteness can contribute to the presence of weak factors, as not all sources of risk are adequately spanned across all assets.
- ...when constructing a portfolio that is neutral to systematic risk but instead only exposed to unsystematic risk (see DelloPreite, Raponi, and Zaffaroni 2025).

# Key insight of this paper

- Back to the famous equation,

$$\mu(\text{rewards}) = B(\text{risk}) \times \gamma(\text{rewards per unit risk})$$

  - We do not know $B$ but estimate $\widehat{B} = B + er$ (er=estimation error)

# Key insight of this paper

- Back to the famous equation,

$$\mu(\text{rewards}) = B(\text{risk}) \times \gamma(\text{rewards per unit risk})$$

  - We do not know $B$ but estimate $\widehat{B} = B + er$ (er=estimation error)
- Taxonomy of asset pricing econometrics

| | Small $T$ | Large $T$ |
|---|---|---|
| Strong Factors | $B_{strong} \sim er$ | $B_{strong} \gg er$ |
| Weak Factors | $B_{weak} \ll er$ | $B_{weak} \sim er$ |

# Key insight of this paper

- Back to the famous equation,

$$\mu(\text{rewards}) = B(\text{risk}) \times \gamma(\text{rewards per unit risk})$$

  - We do not know $B$ but estimate $\widehat{B} = B + er$ (er=estimation error)

- Taxonomy of asset pricing econometrics

| | Small $T$ | Large $T$ |
|---|---|---|
| Strong Factors | $B_{strong} \sim er$ | $B_{strong} \gg er$ |
| Weak Factors | $B_{weak} \ll er$ | $B_{weak} \sim er$ |

1. Traditionally, estimation errors in estimated beta are cause of trouble
2. We flip it as a blessing to reveal whether a given factor is weak or not:
3. Key insight I: distinguish between behaviour of (sum of) $B^2$ from (sum of) $|B|$ -
4. Key insight II: power of test from estimating zero-beta rate (intercept) instead of risk premia (slope).

# Contribution to the literature

- The literature has focused on the issue of weak factors under large $T$ setup

# Contribution to the literature

- The literature has focused on the issue of weak factors under large $T$ setup
  - How to overcome very Weak (Spurious) Factors in testing Strong Factors
    - Kan and Zhang (1999), Kleibergen (2009), Bryzgalova (2016)

# Contribution to the literature

- The literature has focused on the issue of weak factors under large $T$ setup
  - How to overcome very Weak (Spurious) Factors in testing Strong Factors
    - Kan and Zhang (1999), Kleibergen (2009), Bryzgalova (2016)
  - How to identify Price of Risk in Weak Factors
    - Giglio, Xiu and Zhang (2021), Lettau and Pelger (2020), Anatolyev and Mikusheva (2022), Kleibergen and Zhan (2023)

# Contribution to the literature

- The literature has focused on the issue of weak factors under large $T$ setup
  - How to overcome very Weak (Spurious) Factors in testing Strong Factors
    - Kan and Zhang (1999), Kleibergen (2009), Bryzgalova (2016)
  - How to identify Price of Risk in Weak Factors
    - Giglio, Xiu and Zhang (2021), Lettau and Pelger (2020), Anatolyev and Mikusheva (2022), Kleibergen and Zhan (2023)
  - How to test whether a factor of interest is Weak/Semi-strong
    - Pesaran (2012), Pesaran and Smith (2021), Connor and Korajczyk (2022)

# Contribution to the literature

- The literature has focused on the issue of weak factors under large $T$ setup
  - How to overcome very Weak (Spurious) Factors in testing Strong Factors
    - Kan and Zhang (1999), Kleibergen (2009), Bryzgalova (2016)
  - How to identify Price of Risk in Weak Factors
    - Giglio, Xiu and Zhang (2021), Lettau and Pelger (2020), Anatolyev and Mikusheva (2022), Kleibergen and Zhan (2023)
  - How to test whether a factor of interest is Weak/Semi-strong
    - Pesaran (2012), Pesaran and Smith (2021), Connor and Korajczyk (2022)

# Our Contribution

- We propose a novel test for weak factors under *fixed $T$* and *large $N$* setup that:
  - builds on the two-pass methodology - simple and intuitive.
  - handles conditional asset pricing models but robust to
    - misspecified conditional dynamics (semi-parametric)
    - omitted risk factors (PCA)
  - detect whether observed risk factors are *(locally)* weak or not

# Outline

Overview

Economy

- Conditional Asset Pricing Set-Up
- Benchmark Case: No Strong Factor
- Observed Strong Factors
- Unobserved Strong Factors

Simulation

Empirical Application

Conclusion

# Outline

Overview

Economy

Conditional Asset Pricing Set-Up

Benchmark Case: No Strong Factor

Observed Strong Factors

Unobserved Strong Factors

Simulation

Empirical Application

Conclusion

- Conditional Factor Structure for asset $i = 1, \cdots, N$ at $t = 1, \cdots, T$:

$$R_{it} = \alpha_{it-1} + \underbrace{\boldsymbol{\beta}'_{fit-1}\boldsymbol{f}_t}_{\text{strong}} + \underbrace{\boldsymbol{\beta}'_{git-1}\boldsymbol{g}_t}_{\text{weak}} + e_{it},$$

where

$$\boldsymbol{\beta}_{fit-1} = \left(\beta_{f_1 it-1}, \cdots, \beta_{f_K it-1}\right)', \ \boldsymbol{f}_t = \left(f_{1t}, \cdots, f_{Kt}\right)$$
$$\boldsymbol{\beta}_{git-1} = \left(\beta_{g_1 it-1}, \cdots, \beta_{g_L it-1}\right)', \ \boldsymbol{g}_t = \left(g_{1t}, \cdots, g_{Lt}\right)$$

# RGP+APT+Local Smoothness

- We can treat (smoothness assumption) the conditional model as a locally unconditional model:

$$\mathbf{R}_t = \gamma_{zt-1}\mathbf{1}_N + \mathbf{B}_f\boldsymbol{\delta}_{ft} + \mathbf{B}_g\boldsymbol{\delta}_{gt} + \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{\delta}_{ft}$ and $\boldsymbol{\delta}_{gt}$ are *expost* risk premia:

$$\boldsymbol{\delta}_{ft} = \boldsymbol{\gamma}_{ft-1} + \boldsymbol{f}_t - E\left[\boldsymbol{f}_t|\mathcal{I}_{t-1}\right], \boldsymbol{\delta}_{gt} = \boldsymbol{\gamma}_{gt-1} + \boldsymbol{g}_t - E\left[\boldsymbol{g}_t|\mathcal{I}_{t-1}\right]$$

# Local Factor Strength

- For some $0 \leq \rho \leq 1$, the matrix $\mathbf{B}_g$ satisfies

$$\|\mathbf{B}_g\|^2 \asymp O\left(N^{\rho}\right), \ \|\mathbf{B}_g' 1_N\| \asymp o\left(N^{\frac{\rho+1}{2}}\right)$$

# Local Factor Strength

- For some $0 \leq \rho \leq 1$, the matrix $\mathbf{B}_g$ satisfies

$$\|\mathbf{B}_g\|^2 \asymp O\left(N^{\rho}\right), \ \|\mathbf{B}_g' 1_N\| \asymp o\left(N^{\frac{\rho+1}{2}}\right)$$

  - When $\rho = 1$, $\frac{\mathbf{B}_g' \mathbf{B}_g}{N} \asymp O(1)$, or $\boldsymbol{g}_t$ is strong

# Local Factor Strength

- For some $0 \leq \rho \leq 1$, the matrix $\mathbf{B}_g$ satisfies

$$\|\mathbf{B}_g\|^2 \asymp O\left(N^{\rho}\right), \ \|\mathbf{B}_g' 1_N\| \asymp o\left(N^{\frac{\rho+1}{2}}\right)$$

  - When $\rho = 1$, $\frac{\mathbf{B}_g' \mathbf{B}_g}{N} \asymp O(1)$, or $\boldsymbol{g}_t$ is strong
  - When $\rho = 0$, $\mathbf{B}_g' \mathbf{B}_g \asymp O(1)$, as $\boldsymbol{g}_t$ is weak but $\|\mathbf{B}_g' 1_N\| \asymp o\left(N^{\frac{1}{2}}\right)$ (diverges).

# Local Factor Strength

- For some $0 \leq \rho \leq 1$, the matrix $\mathbf{B}_g$ satisfies

$$\|\mathbf{B}_g\|^2 \asymp O\left(N^{\rho}\right), \ \|\mathbf{B}_g' 1_N\| \asymp o\left(N^{\frac{\rho+1}{2}}\right)$$

  - When $\rho = 1$, $\frac{\mathbf{B}_g' \mathbf{B}_g}{N} \asymp O(1)$, or $\boldsymbol{g}_t$ is strong
  - When $\rho = 0$, $\mathbf{B}_g' \mathbf{B}_g \asymp O(1)$, as $\boldsymbol{g}_t$ is weak but $\|\mathbf{B}_g' 1_N\| \asymp o\left(N^{\frac{1}{2}}\right)$ (diverges).
  - The difference in the convergence speed plays a key role to learn $\rho$
    - Analogy to well-spread portfolio $\boldsymbol{w}$, $\boldsymbol{w}' \mathbf{1}_N = 1$ and $\boldsymbol{w}'\boldsymbol{w} \to 0$

# Outline

Overview

Economy

- Conditional Asset Pricing Set-Up
- Benchmark Case: No Strong Factor
- Observed Strong Factors
- Unobserved Strong Factors

Simulation

Empirical Application

Conclusion

# Target Equation

- First, we consider the case that there is no factor $\boldsymbol{f}$:
  - RGP

$$\mathbf{R}_t = \boldsymbol{\alpha}_{t-1} + \mathbf{B}_g \boldsymbol{g}_t + \boldsymbol{\epsilon}_t$$

# Target Equation

- First, we consider the case that there is no factor $\boldsymbol{f}$:
  - RGP

$$\mathbf{R}_t = \boldsymbol{\alpha}_{t-1} + \mathbf{B}_g \boldsymbol{g}_t + \boldsymbol{\epsilon}_t$$

  - Along with the pricing, $\mu = B \times \gamma$

$$\mathbf{R}_t = \gamma_{zt-1} \mathbf{1}_N + \mathbf{B}_g \boldsymbol{\delta}_{gt} + \boldsymbol{\epsilon}_t,$$

    which gives the target equation:

$$\overline{\mathbf{R}} = \overline{\gamma}_z \mathbf{1}_N + \mathbf{B}_g \overline{\boldsymbol{\delta}}_g + \overline{\boldsymbol{\epsilon}}$$

  - Note that we are interested in whether $\boldsymbol{g}$ is weak or not

# FMB two-pass

- First-pass time-series OLS gives

$$\widehat{\mathbf{B}}_{\mathrm{g0}} = \mathbf{B}_g + \epsilon \mathcal{P}_g,$$

where $\mathbf{R} = (\mathbf{R}_1, \cdots, \mathbf{R}_T)'$, $G = (\boldsymbol{g}_1, \cdots, \boldsymbol{g}_T)'$, $\mathcal{J}_T = I_T - \frac{1}{T}\mathbf{1}_T\mathbf{1}_T'$, $\mathcal{P}_g = \mathcal{J}_T G \left(G'\mathcal{J}_T G\right)^{-1}$

# FMB two-pass

- First-pass time-series OLS gives

$$\widehat{\mathbf{B}}_{\text{g0}} = \mathbf{B}_g + \epsilon \mathcal{P}_g,$$

where $\mathbf{R} = (\mathbf{R}_1, \cdots, \mathbf{R}_T)'$, $G = (\boldsymbol{g}_1, \cdots, \boldsymbol{g}_T)'$, $\mathcal{J}_T = I_T - \frac{1}{T}\mathbf{1}_T\mathbf{1}_T'$, $\mathcal{P}_g = \mathcal{J}_T G \left(G'\mathcal{J}_T G\right)^{-1}$

- Second-pass cross-sectional OLS gives:

$$\begin{aligned}\widehat{\mathbf{\Gamma}}_{\text{g0}} &= \begin{bmatrix} \widehat{\gamma}_{0\text{g0}} \\ \widehat{\boldsymbol{\delta}}_{\text{g0}} \end{bmatrix} = \left(\widehat{\mathbf{X}}_{\text{g0}}'\widehat{\mathbf{X}}_{\text{g0}}\right)^{-1}\widehat{\mathbf{X}}_{\text{g0}}'\overline{\mathbf{R}} \\ &\asymp \begin{bmatrix} \overline{\gamma}_z \\ 0_L \end{bmatrix} + \begin{bmatrix} O\left(\frac{\mathbf{B}_g' 1_N}{N}\right) \\ O\left(\frac{\mathbf{B}_g'\mathbf{B}_g}{N}\right) \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N}}\right),\end{aligned}$$

where

$$\widehat{\mathbf{X}}_{\text{g0}} = \left[1_N \ \hat{\mathbf{B}}_{\text{g0}}\right]$$

# FMB two-pass

- First-pass time-series OLS gives

$$\widehat{\mathbf{B}}_{\text{g0}} = \mathbf{B}_g + \epsilon \mathcal{P}_g,$$

where $\mathbf{R} = (\mathbf{R}_1, \cdots, \mathbf{R}_T)'$, $G = (\boldsymbol{g}_1, \cdots, \boldsymbol{g}_T)'$, $\mathcal{J}_T = I_T - \frac{1}{T}\mathbf{1}_T\mathbf{1}_T'$, $\mathcal{P}_g = \mathcal{J}_T G \left(G'\mathcal{J}_T G\right)^{-1}$

- Second-pass cross-sectional OLS gives:

$$\begin{aligned}\widehat{\mathbf{\Gamma}}_{\text{g0}} &= \begin{bmatrix} \widehat{\gamma}_{z0} \\ \widehat{\boldsymbol{\delta}}_{\text{g0}} \end{bmatrix} = \left(\widehat{\mathbf{X}}_{\text{g0}}'\widehat{\mathbf{X}}_{\text{g0}}\right)^{-1}\widehat{\mathbf{X}}_{\text{g0}}'\overline{\mathbf{R}} \\ &\asymp \begin{bmatrix} \overline{\gamma}_0 \\ 0_L \end{bmatrix} + \begin{bmatrix} O\left(\frac{\mathbf{B}_g' 1_N}{N}\right) \\ O\left(\frac{\mathbf{B}_g'\mathbf{B}_g}{N}\right) \end{bmatrix} + O_p\left(\frac{1}{\sqrt{N}}\right),\end{aligned}$$

where

$$\widehat{\mathbf{X}}_{\text{g0}} = \left[1_N \ \hat{\mathbf{B}}_{\text{g0}}\right]$$

# Properties of FMB 1 - Risk Premia (slope)

**Theorem 1.** *Under some Assumptions, the two-pass estimator* $\widehat{\boldsymbol{\delta}}_{g0}$ *in* $\widehat{\boldsymbol{\Gamma}}_{g0} = \left[\widehat{\gamma}_{z0}\ \widehat{\boldsymbol{\delta}}'_{g0}\right]'$ *behaves as follows:*

| | $\widehat{\boldsymbol{\delta}}_{g0} \to_p$ | $\sqrt{N}\widehat{\boldsymbol{\delta}}_{g0} \to_d$ |
|---|---|---|
| $\rho < \frac{1}{2}$ | $\mathbf{0}_L$ | $\mathcal{N}\left(\mathbf{0}_L, \frac{\kappa_4 + T s_4}{T^2 s_2^2} G' \mathcal{J}_T G\right)$ |
| $\rho = \frac{1}{2}$ | | $\mathcal{N}\left(\mathbf{0}_L, \frac{\kappa_4 + T s_4}{T^2 s_2^2} G' \mathcal{J}_T G\right) + O_p(1)$ |
| $\frac{1}{2} < \rho < 1$ | $\widehat{\boldsymbol{\delta}}_{g0} \nrightarrow_p \overline{\boldsymbol{\delta}}_g$ | $\pm\infty$ |
| $\rho = 1$ | | |

*where* $s_2 = \lim_N \frac{1}{N}\sum_i \epsilon_{it}^2$, $\kappa_4 = \left(\lim_N \frac{1}{N}\sum_i \epsilon_{it}^4 - 3s_4\right)$ *and* $s_4 = \lim_N \frac{1}{N}\sum_i E\left[\epsilon_{it}^2\right]^2$

# Relation to Standard OLS

**Theorem 2.** *Under the assumption that residuals are normal i.i.d, the OLS statistics $R^2_{g0}$ and t-stats and F-stat on $\widehat{\boldsymbol{\delta}}_{g0}$ behaves as follows:*

| | $R^2_{g0} \to_p$ | $t_{g0,k} \to_p$ | $F_{g0} \to_p$ |
|---|---|---|---|
| $\rho < \frac{1}{2}$ | | $\mathcal{N}(0,1)$ | $\frac{\chi^2_L}{L}$ |
| $\rho = \frac{1}{2}$ | 0 | $\mathcal{N}(0,1) + O_p(1)$ | $\frac{\chi^2_L}{L} + O_p(1)$ |
| $\frac{1}{2} < \rho < 1$ | | $\pm\infty$ | $\infty$ |
| $\rho = 1$ | $(0,1)$ | | |

# Properties of FMB 2 - Zero-Beta Rate (intercept)

**Theorem 3.** *Under some Assumptions, the two-pass estimator* $\widehat{\gamma}_{z0}$ *in* $\widehat{\mathbf{\Gamma}}_{g0} = \left[\widehat{\gamma}_{z0}\ \widehat{\boldsymbol{\delta}}'_{g0}\right]'$ *behaves as follows:*

| | $\widehat{\gamma}_{z0} \to_p$ | $\sqrt{N}\left(\widehat{\gamma}_{0g0} - \overline{\gamma}_0\right) \to_d$ |
|---|---|---|
| $\rho = 0$ | | $\mathcal{N}\left(0, \frac{s_2}{T}\right)$ |
| $0 < \rho < 1$ | $\overline{\gamma}_z$ | $\pm\infty$ |
| $\rho = 1$ | $\widehat{\gamma}_{z0} \nrightarrow_p \overline{\gamma}_z$ | |

*where* $s_2 = \lim_N \frac{1}{N} \sum_i \epsilon_{it}^2$

# Properties of FMB 2 - Zero-Beta Rate (intercept)

**Theorem 3.** *Under some Assumptions, the two-pass estimator* $\widehat{\gamma}_{z0}$ *in* $\widehat{\mathbf{\Gamma}}_{g0} = \left[\widehat{\gamma}_{z0}\ \widehat{\boldsymbol{\delta}}'_{g0}\right]'$ *behaves as follows:*

| | $\widehat{\gamma}_{z0} \to_p$ | $\sqrt{N}\left(\widehat{\gamma}_{0g0} - \overline{\gamma}_0\right) \to_d$ |
|---|---|---|
| $\rho = 0$ | | $\mathcal{N}\left(0, \frac{s_2}{T}\right)$ |
| $0 < \rho < 1$ | $\overline{\gamma}_z$ | $\pm\infty$ |
| $\rho = 1$ | $\widehat{\gamma}_{z0} \nrightarrow_p \overline{\gamma}_z$ | |

*where* $s_2 = \lim_N \frac{1}{N}\sum_i \epsilon_{it}^2$

- *Given that we do not observe* $\overline{\gamma}_z$ *(except* $R$ *is an excess return), the asymptotic distribution is not directly useful*
  - *This property originates from that* $\widehat{\gamma}_{z0}$ *contains* $\frac{\mathbf{1}'_N \mathbf{B}_g}{N}$
- *Hence, we propose a new test using* $\sqrt{N}\frac{\mathbf{1}'_N \widehat{\mathbf{B}}_{g0}}{N}\overline{\boldsymbol{\delta}}_g$

# Weakness Test

**Theorem 4.** *Under some Assumptions,* $\sqrt{N}\frac{\mathbf{1}_N'\widehat{\mathbf{B}}_{g0}}{N}\overline{\boldsymbol{\delta}}_g$ *behaves as follows:*

| | $\sqrt{N}\frac{\mathbf{1}_N'\widehat{\mathbf{B}}_{g0}}{N}\overline{\boldsymbol{\delta}}_g \to_d$ |
|---|---|
| $\rho = 0$ | $\mathcal{N}\left(0, s_2\overline{\boldsymbol{\delta}}_g'\left(G'\mathcal{J}_T G\right)^{-1}\overline{\boldsymbol{\delta}}_g\right)$ |
| $0 < \rho < 1$ | $\pm\infty$ |
| $\rho = 1$ | |

- *Furthermore, we observe all the elements for the asymptotic variance except* $s_2 = \lim_N \frac{1}{N}\sum_i \epsilon_{it}^2$*!*

# Estimation of Asymptotic Variance

- Recall that when $\rho < \frac{1}{2}$
  - $\sqrt{N}\widehat{\boldsymbol{\delta}}_{\text{g0}} \to_d \mathcal{N}\left(\mathbf{0}_L, \frac{\kappa_4 + T s_4}{T^2 s_2^2} G' \mathcal{J}_T G\right)$

# Estimation of Asymptotic Variance

- Recall that when $\rho < \frac{1}{2}$
  - $\sqrt{N}\widehat{\boldsymbol{\delta}}_{\text{g0}} \to_d \mathcal{N}\left(\mathbf{0}_L, \frac{\kappa_4 + T s_4}{T^2 s_2^2} G' \mathcal{J}_T G\right)$
- We need to estimate $s_2 = \lim_N \frac{1}{N} \sum_i \epsilon_{it}^2$, $\kappa_4 = \left(\lim_N \frac{1}{N} \sum_i \epsilon_{it}^4 - 3 s_4\right)$ and $s_4 = \lim_N \frac{1}{N} \sum_i E\left[\epsilon_{it}^2\right]^2$ !
  - We can do that by exploiting estimated residuals from first-pass as well as those from second-pass

# Summary of tests

- We utilize two tests: (i) coefficients on the noisy betas from FMB and (ii) average of the noisy betas

| | $\sqrt{N}\widehat{\boldsymbol{\delta}}_{\text{g0}}$ | $\frac{\mathbf{1}_N'\widehat{\mathbf{B}}_{\text{g0}}}{\sqrt{N}}\overline{\boldsymbol{\delta}}_g$ |
|---|---|---|
| $\rho = 0$ | Null | Null |
| $0 < \rho < \frac{1}{2}$ | | Alternative |
| $\rho \geq \frac{1}{2}$ | Alternative | |

# Outline

Overview

Economy

- Conditional Asset Pricing Set-Up
- Benchmark Case: No Strong Factor
- Observed Strong Factors
- Unobserved Strong Factors

Simulation

Empirical Application

Conclusion

# Modified FMB two-pass

- First-pass time-series OLS gives

$$\widehat{\mathbf{B}}_f = \mathbf{B}_f + \epsilon\mathcal{P}_f, \widehat{\mathbf{B}}_g = \mathbf{B}_g + \epsilon\mathcal{P}_{g_\perp},$$

where $\mathcal{P}_f = \mathcal{J}_T F \left(F'\mathcal{J}_T F\right)^{-1}$, $\mathcal{P}_{g\perp} = \mathcal{J}_T G_\perp \left(G_\perp'\mathcal{J}_T G_\perp\right)^{-1}$

# Modified FMB two-pass

- First-pass time-series OLS gives

$$\widehat{\mathbf{B}}_f = \mathbf{B}_f + \epsilon\mathcal{P}_f, \widehat{\mathbf{B}}_g = \mathbf{B}_g + \epsilon\mathcal{P}_{g_\perp},$$

  where $\mathcal{P}_f = \mathcal{J}_T F \left(F'\mathcal{J}_T F\right)^{-1}$, $\mathcal{P}_{g\perp} = \mathcal{J}_T G_\perp \left(G_\perp'\mathcal{J}_T G_\perp\right)^{-1}$
- Second-pass cross-sectional OLS gives:

$$\widehat{\mathbf{\Gamma}}_g = \begin{bmatrix} \widehat{\gamma}_z \\ \widehat{\boldsymbol{\delta}}_g \end{bmatrix} = \left(\widehat{\mathbf{X}}_g'\widehat{\mathbf{X}}_g\right)^{-1}\widehat{\mathbf{X}}_g'\left(\overline{\mathbf{R}} - \widehat{\mathbf{B}}_f\overline{\boldsymbol{\delta}}_f\right),$$

  where

$$\widehat{\mathbf{X}}_g = \left[1_N\ \widehat{\mathbf{B}}_g\right]$$

# Modified FMB two-pass

- First-pass time-series OLS gives

$$\widehat{\mathbf{B}}_f = \mathbf{B}_f + \epsilon\mathcal{P}_f, \widehat{\mathbf{B}}_g = \mathbf{B}_g + \epsilon\mathcal{P}_{g_\perp},$$

  where $\mathcal{P}_f = \mathcal{J}_T F \left(F'\mathcal{J}_T F\right)^{-1}$, $\mathcal{P}_{g\perp} = \mathcal{J}_T G_\perp \left(G_\perp'\mathcal{J}_T G_\perp\right)^{-1}$
- Second-pass cross-sectional OLS gives:

$$\widehat{\mathbf{\Gamma}}_g = \begin{bmatrix} \widehat{\gamma}_z \\ \widehat{\boldsymbol{\delta}}_g \end{bmatrix} = \left(\widehat{\mathbf{X}}_g'\widehat{\mathbf{X}}_g\right)^{-1}\widehat{\mathbf{X}}_g'\left(\overline{\mathbf{R}} - \widehat{\mathbf{B}}_f\overline{\boldsymbol{\delta}}_f\right),$$

  where

$$\widehat{\mathbf{X}}_g = \left[1_N\ \widehat{\mathbf{B}}_g\right]$$

- If we include $\widehat{\mathbf{B}}_f$ in the second pass regressor
  - It is well known that the estimator is biased due to estimation error
  - The bias-correction such as Shanken (1992) does not work

# Slight modification of tests

- Two tests have similar properties

| | $\sqrt{N}\widehat{\boldsymbol{\delta}}_g \to_d$ | $\frac{\mathbf{1}_N'\widehat{\mathbf{B}}_g}{\sqrt{N}}\overline{\boldsymbol{\delta}}_g \to_d$ |
|---|---|---|
| $\rho = 0$ | $\mathcal{N}(\mathbf{0}_L, V_1)$ | $\mathcal{N}(0, V_2)$ |
| $\rho < \frac{1}{2}$ | | |
| $\rho = \frac{1}{2}$ | $\mathcal{N}(\mathbf{0}_L, V_1) + O_p(1)$ | $\pm\infty$ |
| $\frac{1}{2} < \rho \leq 1$ | $\pm\infty$ | |

where

$$V_1 = \frac{s_4}{s_2^2}\mathbf{I}'\mathbf{I}G_\perp' G_\perp + \frac{\kappa_4}{s_2^2} G_\perp' \text{diag}(\mathbf{I} \odot \mathbf{I}) G_\perp$$

$$\mathbf{I} = \frac{1_T}{T} - \mathcal{P}_f \overline{\boldsymbol{\delta}}_f$$

$$V_2 = s_2 \overline{\boldsymbol{\delta}}_g' (G_\perp' G_\perp)^{-1} \overline{\boldsymbol{\delta}}_g$$

- Furthermore, we can make the tests feasible using consistent estimators for components in the asymptotic variance

# Outline

Overview

Economy

- Conditional Asset Pricing Set-Up
- Benchmark Case: No Strong Factor
- Observed Strong Factors
- Unobserved Strong Factors

Simulation

Empirical Application

Conclusion

# PCA

- We borrow the idea of Giglio and Xiu (2021) and span the (strong) factors' space by PCA.
- Following the local PCA methodology of Zaffaroni (2025), we obtain the systematic factors up to rotation (when $T$ fixed)

$$F_* - F\tilde{H} \to_p 0_{T \times K}$$

# Modified FMB two-pass with PCA factors

- First-pass time-series OLS gives

$$\widehat{\mathbf{B}}_{f_*} = \mathbf{B}_{f_*} + \epsilon_* \mathcal{P}_{f_*},$$
$$\widehat{\mathbf{B}}_{g_*} = \mathbf{B}_{g_*} + \epsilon_* \mathcal{P}_{g_{*\perp}},$$

where $\mathcal{P}_{f_*} = \mathcal{J}_T F_* \left(F_*' \mathcal{J}_T F_*\right)^{-1}$, $\mathcal{P}_{g_{*\perp}} = \mathcal{J}_T G_{*\perp} \left(G_{*\perp}' \mathcal{J}_T G_{*\perp}\right)^{-1}$

# Modified FMB two-pass with PCA factors

- First-pass time-series OLS gives

$$\widehat{\mathbf{B}}_{f_*} = \mathbf{B}_{f_*} + \epsilon_* \mathcal{P}_{f_*},$$
$$\widehat{\mathbf{B}}_{g_*} = \mathbf{B}_{g_*} + \epsilon_* \mathcal{P}_{g_{*\perp}},$$

  where $\mathcal{P}_{f_*} = \mathcal{J}_T F_* \left(F_*' \mathcal{J}_T F_*\right)^{-1}$, $\mathcal{P}_{g_{*\perp}} = \mathcal{J}_T G_{*\perp} \left(G_{*\perp}' \mathcal{J}_T G_{*\perp}\right)^{-1}$

- Second-pass cross-sectional OLS gives:

$$\widehat{\mathbf{\Gamma}}_{g_*} = \begin{bmatrix} \widehat{\gamma}_{z_*} \\ \widehat{\boldsymbol{\delta}}_{g_*} \end{bmatrix} = \left(\widehat{\mathbf{X}}_{g_*}' \widehat{\mathbf{X}}_{g_*}\right)^{-1} \widehat{\mathbf{X}}_{g_*}' \left(\overline{\mathbf{R}} - \widehat{\mathbf{B}}_{f_*} \overline{\boldsymbol{\delta}}_{f_*}\right),$$

  where

$$\widehat{\mathbf{X}}_{g_*} = \left[1_N \ \widehat{\mathbf{B}}_{g_*}\right]$$

# Slight modification of tests

- Two tests have similar properties

| | $\sqrt{N}\widehat{\boldsymbol{\delta}}_{g_*} \to_d$ | $\frac{\mathbf{1}_N'\widehat{\mathbf{B}}_{g_*}}{\sqrt{N}}\overline{\boldsymbol{\delta}}_{g_*} \to_d$ |
|---|---|---|
| $\rho = 0$ | $\mathcal{N}(\mathbf{0}_L, V_{1*})$ | $\mathcal{N}(0, V_{2*})$ |
| $\rho < \frac{1}{2}$ | | |
| $\rho = \frac{1}{2}$ | $\mathcal{N}(\mathbf{0}_L, V_{1*}) + O_p(1)$ | $\pm\infty$ |
| $\frac{1}{2} < \rho \leq 1$ | $\pm\infty$ | |

where

$$V_{1*} = \frac{s_4}{s_2^2}\mathbf{l}'\mathbf{l}G_\perp'G_\perp + \frac{\kappa_4}{s_2^2}G_\perp'\text{diag}(\mathbf{l}\odot\mathbf{l})G_\perp$$

$$V_{2*} = c_* + s_2\overline{\boldsymbol{\delta}}_g'(G_\perp'G_\perp)^{-1}\overline{\boldsymbol{\delta}}_g$$

- Furthermore, we can make the tests feasible using consistent estimator of the corresponding asymptotic variances

# Outline

Overview

Economy

- Conditional Asset Pricing Set-Up
- Benchmark Case: No Strong Factor
- Observed Strong Factors
- Unobserved Strong Factors

Simulation

Empirical Application

Conclusion

# Simulation Design I (conventional MC)

1. Calibration: MacKinlay and Pastor (2000)

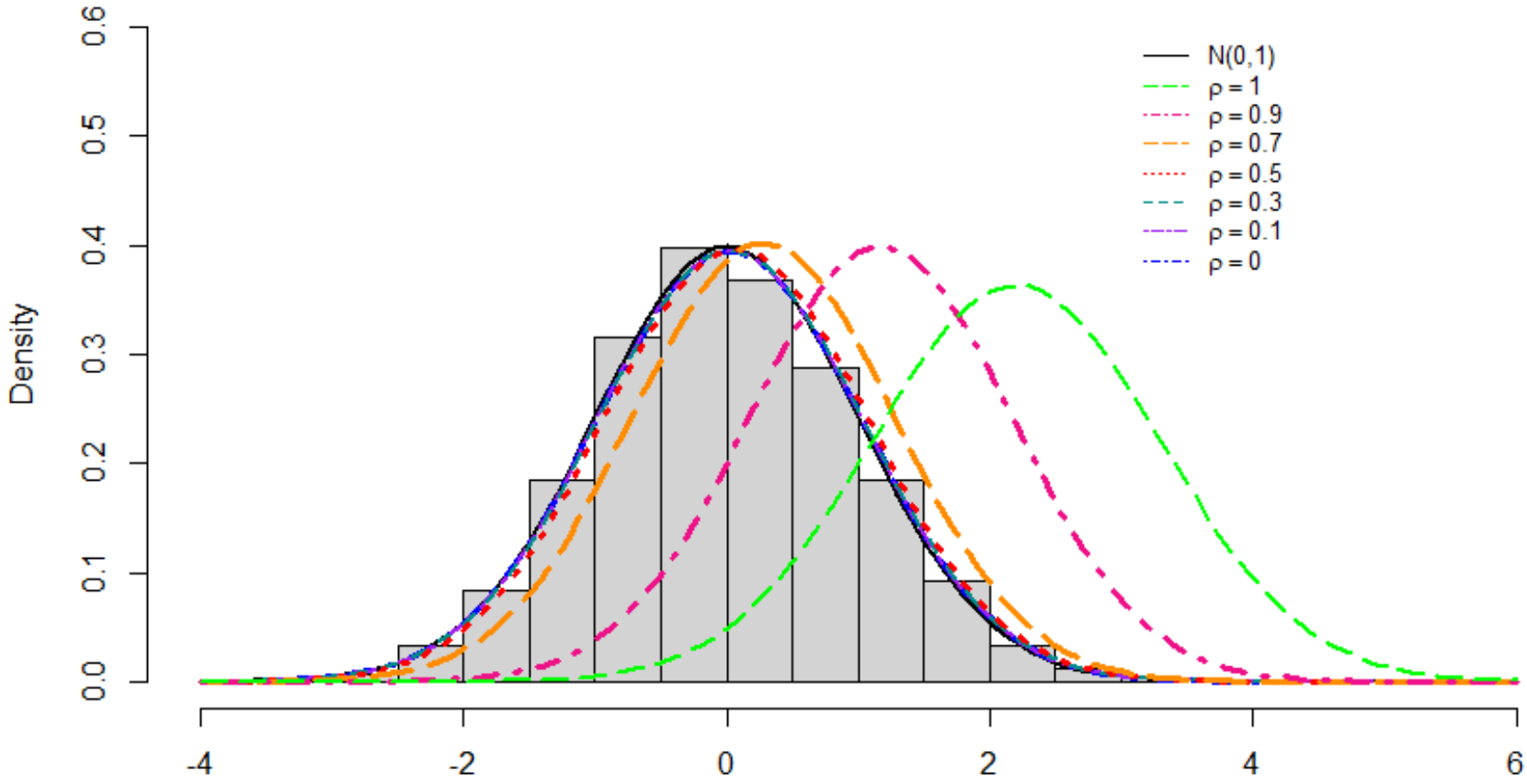
$$R_{it} = 0 + \boldsymbol{\beta}_{fi}\boldsymbol{f}_t + \boldsymbol{\beta}_{gi}\boldsymbol{g}_t + e_{it}$$

2. We consider a single strong factor and a single weak factor, $N = 3000,\ T = 24$
3. We focus on the distribution of the following two tests

| | test 1: $\frac{1}{\sqrt{\widehat{AsyVar}}}\left(\sqrt{N}\widehat{\boldsymbol{\delta}}_g\right)$ | test 2: $\frac{1}{\sqrt{\widehat{AsyVar}}}\left(\frac{\mathbf{1}_N'\widehat{\mathbf{B}}_g}{\sqrt{N}}\overline{\boldsymbol{\delta}}_g\right)$ |
|---|---|---|
| $\rho = 0$ | $\mathcal{N}(0,1)$ | $\mathcal{N}(0,1)$ |
| $\rho < \frac{1}{2}$ | | |
| $\rho = \frac{1}{2}$ | $\mathcal{N}(0,1) + O_p(1)$ | $\pm\infty$ |
| $\frac{1}{2} < \rho \leq 1$ | $\pm\infty$ | |

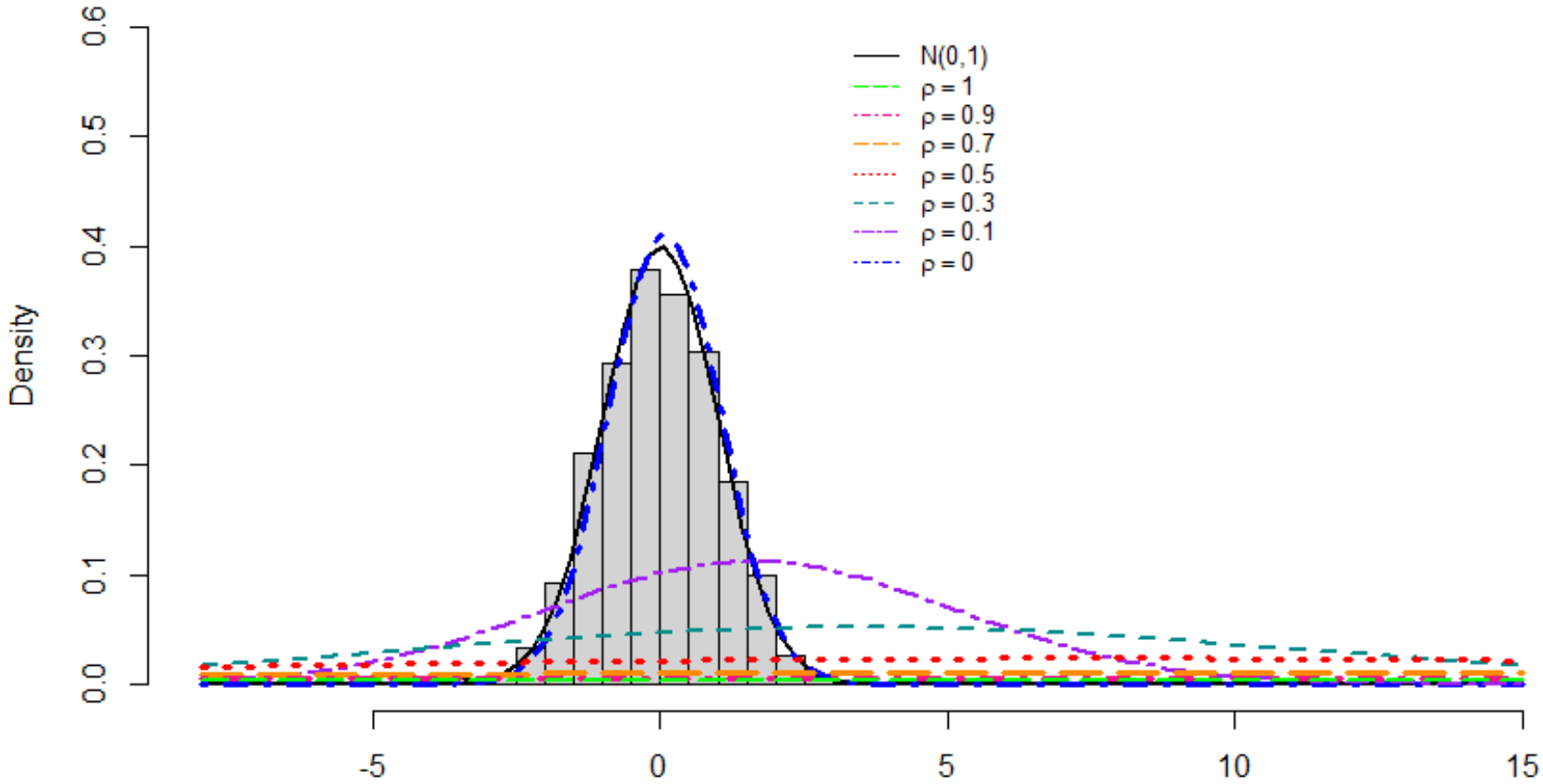
# test 1: under the null $\rho < \frac{1}{2}$ + DGP with $\rho \in [0, 1]$

- 3000 repetitions



# test 2: under the null $\rho = 0$ + DGP with $\rho \in [0, 1]$

- 3000 repetitions



# Simulation Design II (cool MC - SPCA)

- Giglio, Xiu and Zhang (2024) introduced the Supervised PCA (SPCA) to estimate the risk premia of observed weak factors.
- Their two-step methodology:(i) identify the subset of assets where candidate risk factor is strong; (ii) two-pass estimation of risk premia over the subset of assets.
- We demonstrate how our methodology (KRZ) can be used sequentially with SPCA:
  1. Use KRZ to test whether a given candidate risk factor $g_t$ is weak, semi-strong, or strong.
  2. Use PCA (Giglio and Xiu 2020) or SPCA to estimate $g_t$ risk premia.

# KRZ and SPCA in the presence of a WEAK factor

**Panel A: $s_0 = 1, T = 12$**

| | | KRZ PCA test | SPCA + KRZ PCA test | | | KRZ on Non Selected |
|---|---|---|---|---|---|---|
| N | Scenario | % Reject | % Reject | $N_{SPCA}$ | % (mode) | % Reject |
| 200, $N_0$=20 | G orth F | 3.2 | 96.8 | 20 | 64.6 | 1.8 |
| | G = F + z | 11.0 | 80.0 | 20 | 59.8 | 1.0 |
| | G ∼ F (cor = 0.99) | 1.6 | 1.8 | 70 | 30.8 | 9.0 |
| 500, $N_0$=50 | G orth F | 8.8 | 96.2 | 50 | 90.8 | 1.5 |
| | G = F + z | 7.4 | 88.8 | 50 | 32.4 | 1.4 |
| | G ∼ F (cor = 0.99) | 1.2 | 1.8 | 400 | 16.2 | 10.3 |
| 1000, $N_0$=100 | G orth F | 11.2 | 100.0 | 100 | 87.2 | 0.8 |
| | G = F + z | 9.4 | 98.8 | 100 | 17.4 | 0.4 |
| | G ∼ F (cor = 0.99) | 0.8 | 5.6 | 150 | 15.8 | 2.4 |

**Panel B: $s_0 = 0.5, T = 12$**

| | | KRZ PCA test | SPCA + KRZ PCA test | | | KRZ on Non Selected |
|---|---|---|---|---|---|---|
| N | Scenario | % Reject | % Reject | $N_{SPCA}$ | % (mode) | % Reject |
| 200, $N_0$=20 | G orth F | 1.2 | 84.0 | 70 | 35.8 | 2.6 |
| | G = F + z | 0.0 | 86.6 | 20 | 48.0 | 0.4 |
| | G ∼ F (cor = 0.99) | 1.2 | 0.6 | 120 | 32.8 | 9.8 |
| 500, $N_0$=50 | G orth F | 1.6 | 81.0 | 50 | 35.6 | 0.7 |
| | G = F +z | 0.8 | 97.2 | 100 | 42.8 | 0.4 |
| | G ∼ F (cor = 0.99) | 2.0 | 2.4 | 150 | 17.0 | 8.4 |
| 1000, $N_0$=100 | G orth F | 1.8 | 99.0 | 100 | 31.8 | 1.0 |
| | G = F + z | 0.4 | 100.0 | 150 | 32.2 | 7.4 |
| | G ∼ F (cor = 0.99) | 0.0 | 10.8 | 150 | 14.6 | 19.4 |

# KRZ and SPCA in the presence of a STRONG factor

**Panel A: $s = 1$, $T = 12$**

| | | KRZ PCA test | SPCA + KRZ PCA test | | | KRZ on Non Selected |
|---|---|---|---|---|---|---|
| N | Scenario | % Reject | % Reject | $N_{SPCA}$ | % (mode) | % Reject |
| 200 | G orth F | 95.2 | 86.0 | 20 | 63.6 | 1.4 |
| | G = F + v | 100.0 | 50.2 | 70 | 48.8 | 6.6 |
| | G ∼ F (cor = 0.99) | 1.0 | 1.6 | 70 | 37.0 | 5.8 |
| 500, $N_0$=50 | G orth F | 99.6 | 94.2 | 50 | 66.8 | 1.0 |
| | G = F + v | 100.0 | 85.8 | 100 | 34.2 | 1.0 |
| | G ∼ F (cor = 0.99) | 2.0 | 4.8 | 100 | 22.6 | 3.6 |
| 1000, $N_0$=100 | G orth F | 99.2 | 97.8 | 150 | 26.2 | 0.6 |
| | G = F + v | 100.0 | 98.4 | 150 | 36.8 | 3.2 |
| | G ∼ F (cor = 0.99) | 10.6 | 7.6 | 150 | 27.8 | 7.2 |

**Panel B: $s = 0.5$, $T = 12$**

| | | KRZ PCA test | SPCA + KRZ PCA test | | | KRZ on Non Selected |
|---|---|---|---|---|---|---|
| N | Scenario | % Reject | % Reject | $N_{SPCA}$ | % (mode) | % Reject |
| 200 | G orth F | 66.6 | 97.4 | 20 | 53.8 | 0.8 |
| | G = F + v | 49.2 | 57.0 | 70 | 44.2 | 9.1 |
| | G ∼ F (cor = 0.99) | 0.2 | 1.2 | 70 | 33.8 | 8.0 |
| 500, $N_0$=50 | G orth F | 78.6 | 99.4 | 50 | 86.6 | 2.2 |
| | G = F + v | 100.0 | 90.4 | 50 | 31.4 | 15.6 |
| | G ∼ F (cor = 0.99) | 4.4 | 1.8 | 400 | 21.8 | 10.8 |
| 1000, $N_0$=100 | G orth F | 44.4 | 100.0 | 100 | 65.6 | 0.4 |
| | G = F + v | 97.2 | 99.0 | 150 | 20.4 | 11.0 |
| | G ∼ F (cor = 0.99) | 0.4 | 4.6 | 150 | 17.0 | 2.8 |

# Outline

Overview

Economy

- Conditional Asset Pricing Set-Up
- Benchmark Case: No Strong Factor
- Observed Strong Factors
- Unobserved Strong Factors

Simulation

Empirical Application

Conclusion

# Outline Empirics

- We focus on the test 2 (zero-beta rate) on the null $\rho = 0$ over 1968-2022
  - Similar message from the test 1 (risk premia) on the null $\rho < \frac{1}{2}$

# Outline Empirics

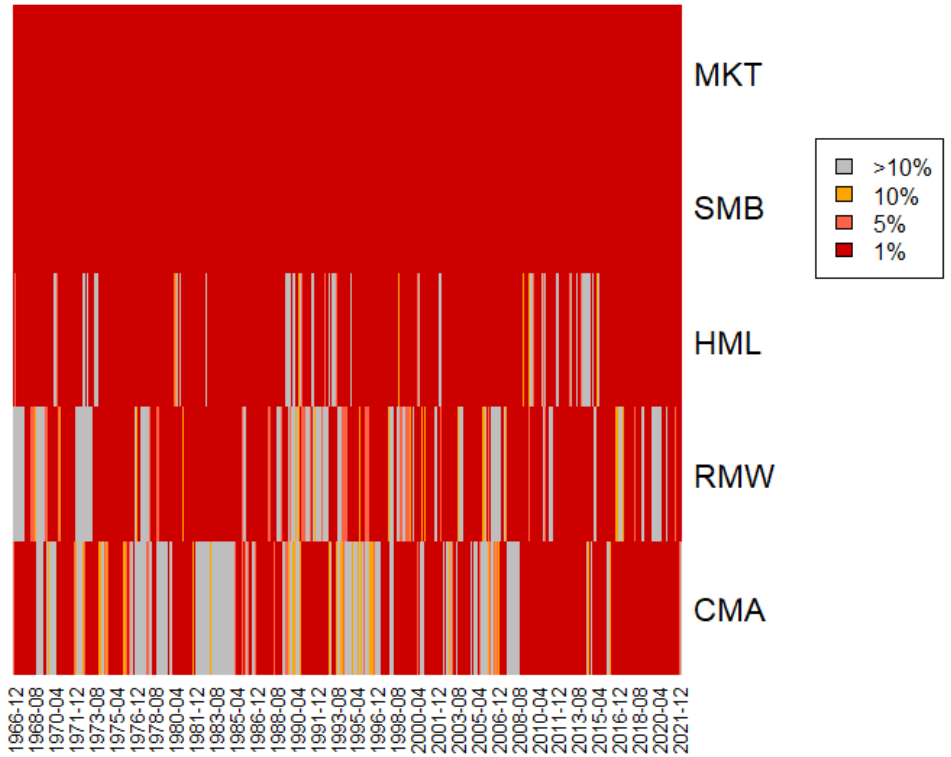
- We focus on the test 2 (zero-beta rate) on the null $\rho = 0$ over 1968-2022
  - Similar message from the test 1 (risk premia) on the null $\rho < \frac{1}{2}$
- Questions:
- Q1 Are there any weak factors in FF5?
  - We test whether a factor in FF5 is weak or not
  - Strong/weak depends on industry

# Outline Empirics

- We focus on the test 2 (zero-beta rate) on the null $\rho = 0$ over 1968-2022
  - Similar message from the test 1 (risk premia) on the null $\rho < \frac{1}{2}$
- Questions:
- Q1 Are there any weak factors in FF5?
  - We test whether a factor in FF5 is weak or not
  - Strong/weak depends on industry
- Q2 Are there any weak factors in the factors zoo?
  - Factor zoo
    - 150 factors from Feng, Giglio and Xiu (2020)
  - Likelihood of being weak on recession/post-publication
- Q3 What is the economic significance of weak factors?
  - Harvesting risk premia of locally strong factor.

# Are there any strong factors in FF5?

- HeatMap (Strong Red - ... - Weak Gray)

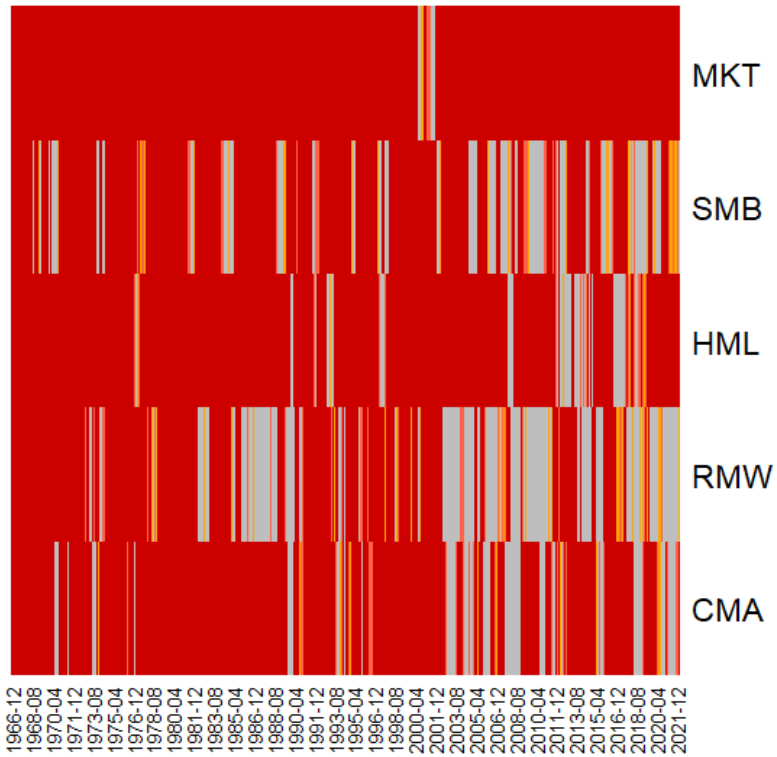



- Null on F and G

| F | G |
|---|---|
| No Strong Factor | MKT |
| CAPM | SMB, HML |
| FF3 | RMW, CMA |

# Stong/Weak of FF5 in Utility Industry

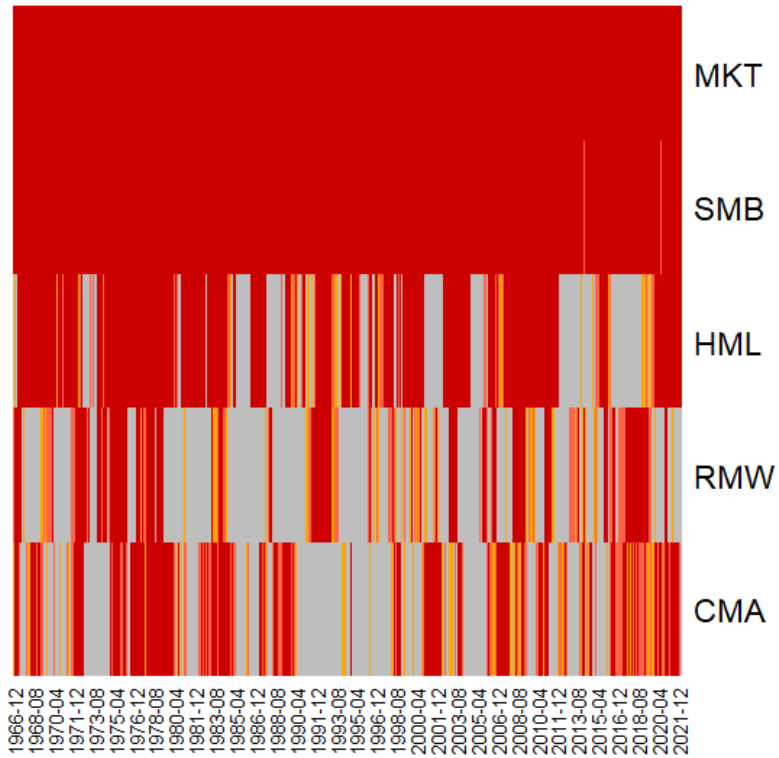
- HeatMap (Strong Red - ... - Weak Gray)




- SMB tend to be weaker in Utility industry
  - 20% of tests in Uitlity vs 0% of tests in CRSP

# Stong/Weak of FF5 in Consumer Nondurables

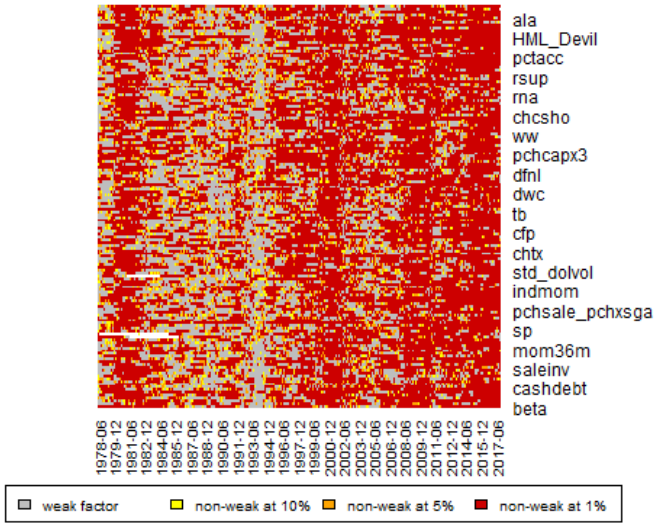
- HeatMap (Strong Red - ... - Weak Gray)




- HML tend to be weaker in Consumer Nondurables industry
  - 27% of tests in Consumer Non-durables vs 10% of tests in CRSP

# Factor Zoo with strong subset of FF5

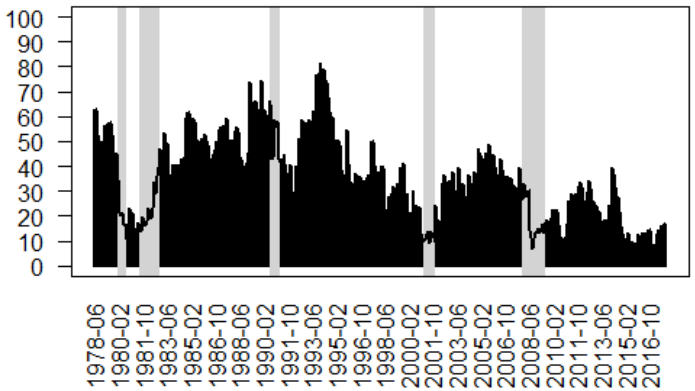
- We select strong factors from FF5 at each local time



# Business Cycle and Weakness of Factors

- Business Cycle and % of Weak factors in factor zoo

$$\% \text{ of weak factors} = a - \underbrace{10.6}_{t=19.69} * \text{NBER recession dummy} + e$$



# Post-publication effect

- What happens to the weakness of a given factor post publication
  - We regress [the dummy on $|t| > 1.96$ from our test] on [the post-publication dummy]

    Strong Dummy using our test $= a + \underbrace{0.19}_{t=45.92} *$Post Publication dummy$+e$

  - Nice contrast with the results that the average returns tend to be lower post publication (McLean and Pontiff, 2016)
    - Public information $=>$ Pervasive & Fair price

# Locally strong - weak factors

- Literature suggests novel methods for how to handle (unconditional) weak factors, where weak factors are defined over a large $T$.
- Our local method (small $T$) can provide novel insights on
  - existence of local weakness?
  - economic significance of weakness?

# Locally strong - weak factors

- Data: Chen-Zimmerman (2022)'s 768 portfolios from 1963:07 to 2023:12
  - Large $T$ PCA risk premia

| PC | mean | std | annual SR |
|---|---|---|---|
| 1 | 0.20 | 1.55 | 0.46 |
| 2 | 0.03 | 0.28 | 0.36 |
| 3 | 0.03 | 0.23 | 0.41 |
| 4 | 0.05 | 0.13 | 1.25 |
| 5 | 0.03 | 0.10 | 1.10 |
| 6 | 0.01 | 0.10 | 0.26 |
| 7 | 0.01 | 0.09 | 0.21 |

    - We consider the fourth/fifth long-term PCs as weak factors
- Our design
  - Set first three (unconditional) PCs as **F**
  - Test whether fourth/fifth (unconditional) PCs **G** are locally weak

# Locally weak/strong of weak factors

- Sorting on the t-stat of the slope (risk premium), one gets

| PCA | Quintile | $\delta_g$ | $g_{t+1}$ |
|---|---|---|---|
| | low $\lvert t\text{-stat}\rvert$ | 0.035 | 0.037 |
| | | 0.042 | 0.053 |
| 4 | | 0.048 | 0.054 |
| | | 0.053 | 0.036 |
| | high $\lvert t\text{-stat}\rvert$ | 0.065 | 0.060 |
| | low $\lvert t\text{-stat}\rvert$ | 0.026 | 0.011 |
| | | 0.024 | 0.032 |
| 5 | | 0.025 | 0.023 |
| | | 0.039 | 0.048 |
| | high $\lvert t\text{-stat}\rvert$ | 0.050 | 0.047 |

  - Unconditionally-weak factors are priced higher when they are locally-strong, i.e., when rejecting the null of locally weak.
  - Unconditionally-weak factors offer higher return next period when they are locally-strong, i.e., when rejecting the null of locally weak.

# Outline

Overview

Economy

- Conditional Asset Pricing Set-Up
- Benchmark Case: No Strong Factor
- Observed Strong Factors
- Unobserved Strong Factors

Simulation

Empirical Application

Conclusion

# Conclusion

- Novel unique methodology to test (local) weakness of factors
- Suitable to conditional asset pricing and robut to misspecified dynamics and risk factors.
  - Asymptotic theory (large $N$ fixed $T$).
  - Simulation evidence I and II (complementary to SPCA).
- Empirical findings
  - anomaly factors: tend to be stronger during recession and post publication
  - unconditionally weak factors tend to be stronger during recession and post publication