

Arbitrage Portfolios  
By  
Soohun Kim, Robert A. Korajczyk, Andreas Neuhierl

---

Nancy R. Xu  
Boston College

July 10, 2019, CICF

# Objective

- ▶ Propose a methodology to estimate arbitrage portfolios by utilizing information contained in firm characteristics for both abnormal returns and factor loadings

# Objective

- ▶ Propose a methodology to estimate arbitrage portfolios by utilizing information contained in firm characteristics for both abnormal returns and factor loadings
- ▶ Enter the long & ongoing debate:

What is the underlying driver of abnormal returns?

# Introduction

- ▶ Long literature: many variables have predictive power of the cross section of asset returns. But it is very difficult to disentangle predictive power due to,

# Introduction

- ▶ Long literature: many variables have predictive power of the cross section of asset returns. But it is very difficult to disentangle predictive power due to,
  1. Beta: exposures to latent or observed systematic factors
  2. Alpha: mispricing
  3. Data snooping: spurious cross-sectional relations due to overfitting

# Introduction

- ▶ Long literature: many variables have predictive power of the cross section of asset returns. But it is very difficult to disentangle predictive power due to,
  1. Beta: exposures to latent or observed systematic factors
  2. Alpha: mispricing
  3. Data snooping: spurious cross-sectional relations due to overfitting
- ▶ One main debate: Risk- or characteristics-base explanations of abnormal returns. Daniel and Titman (1997) develops an influential method of double sorting using (known) risk factor betas and firm characteristics

# Introduction

- ▶ Long literature: many variables have predictive power of the cross section of asset returns. But it is very difficult to disentangle predictive power due to,
  1. Beta: exposures to latent or observed systematic factors
  2. Alpha: mispricing
  3. Data snooping: spurious cross-sectional relations due to overfitting
- ▶ One main debate: Risk- or characteristics-base explanations of abnormal returns. Daniel and Titman (1997) develops an influential method of double sorting using (known) risk factor betas and firm characteristics
- ▶ Some issues:
  1. Regression estimates of systematic risks are known to be relatively imprecise

# Introduction

- ▶ Long literature: many variables have predictive power of the cross section of asset returns. But it is very difficult to disentangle predictive power due to,
  1. Beta: exposures to latent or observed systematic factors
  2. Alpha: mispricing
  3. Data snooping: spurious cross-sectional relations due to overfitting
- ▶ One main debate: Risk- or characteristics-base explanations of abnormal returns. Daniel and Titman (1997) develops an influential method of double sorting using (known) risk factor betas and firm characteristics
- ▶ Some issues:
  1. Regression estimates of systematic risks are known to be relatively imprecise
  2. The methodology handles one characteristics at a time

# Introduction

- ▶ Long literature: many variables have predictive power of the cross section of asset returns. But it is very difficult to disentangle predictive power due to,
  1. Beta: exposures to latent or observed systematic factors
  2. Alpha: mispricing
  3. Data snooping: spurious cross-sectional relations due to overfitting
- ▶ One main debate: Risk- or characteristics-base explanations of abnormal returns. Daniel and Titman (1997) develops an influential method of double sorting using (known) risk factor betas and firm characteristics
- ▶ Some issues:
  1. Regression estimates of systematic risks are known to be relatively imprecise
  2. The methodology handles one characteristics at a time
  3. The sorting by design has implications at the portfolio level (not individual assets)

# Introduction

- ▶ This motivates a new methodology to jointly resolve the three weaknesses in a unified framework: A new generation of technology designed to re-examine the cross-section of asset returns

# Introduction

- ▶ This motivates a new methodology to jointly resolve the three weaknesses in a unified framework: A new generation of technology designed to re-examine the cross-section of asset returns *Feng, Giglio, and Xiu (JF, forthcoming)*; *Kelly, Pruitt, and Su (JFE, forthcoming)*; *Gu, Kelly, and Xiu (WP)*; etc

# Introduction

- ▶ This motivates a new methodology to jointly resolve the three weaknesses in a unified framework: A new generation of technology designed to re-examine the cross-section of asset returns *Feng, Giglio, and Xiu (JF, forthcoming)*; *Kelly, Pruitt, and Su (JFE, forthcoming)*; *Gu, Kelly, and Xiu (WP)*; etc
- ▶ *This paper fits here*

# Main Results

- ▶ Method: Use a variant of Fan et al.'s (2016) "Projected principal component analysis" (PPCA) to estimate a cross section of individual asset returns

# Main Results

- ▶ Method: Use a variant of Fan et al.'s (2016) "Projected principal component analysis" (PPCA) to estimate a cross section of individual asset returns
- ▶ Details: Suppose return generating processes are relatively stable during a period of time and can be expressed in matrices with time-varying factors and a cross-section of alphas and betas,

$$\mathbf{R} = \boldsymbol{\alpha}\mathbf{1}'_T + \mathbf{B}\mathbf{F}' + \mathbf{E}$$

$$\boldsymbol{\alpha} = \mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha$$

$$\mathbf{B} = \mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta$$

⇒  $\mathbf{F}$ : Factors;  $\mathbf{X}$ : a wide range of firm characteristics (+their polynomials)

# Main Results

- ▶ Method: Use a variant of Fan et al.'s (2016) "Projected principal component analysis" (PPCA) to estimate a cross section of individual asset returns
- ▶ Details: Suppose return generating processes are relatively stable during a period of time and can be expressed in matrices with time-varying factors and a cross-section of alphas and betas,

$$\mathbf{R} = \boldsymbol{\alpha} \mathbf{1}'_T + \mathbf{B} \mathbf{F}' + \mathbf{E}$$

$$\boldsymbol{\alpha} = \mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha$$

$$\mathbf{B} = \mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta$$

- ⇒  $\mathbf{F}$ : Factors;  $\mathbf{X}$ : a wide range of firm characteristics (+their polynomials)
- ⇒  $\mathbf{G}_\alpha, \mathbf{G}_\beta$ : a linear function of  $\mathbf{X}$

# Main Results

- ▶ Method: Use a variant of Fan et al.'s (2016) "Projected principal component analysis" (PPCA) to estimate a cross section of individual asset returns
- ▶ Details: Suppose return generating processes are relatively stable during a period of time and can be expressed in matrices with time-varying factors and a cross-section of alphas and betas,

$$\mathbf{R} = \alpha \mathbf{1}'_T + \mathbf{B}\mathbf{F}' + \mathbf{E}$$

$$\alpha = \mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha$$

$$\mathbf{B} = \mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta$$

- ⇒  $\mathbf{F}$ : Factors;  $\mathbf{X}$ : a wide range of firm characteristics (+their polynomials)
- ⇒  $\mathbf{G}_\alpha, \mathbf{G}_\beta$ : a linear function of  $\mathbf{X}$
- ⇒ Estimation: (1) using PPCA to obtain  $\widehat{\mathbf{G}}_\beta$ , (2) estimate  $\widehat{\mathbf{G}}_\alpha$  using regression given  $\widehat{\mathbf{G}}_\beta$ , (3) construct arbitrage portfolios from  $\widehat{\mathbf{G}}_\alpha$

# Main Results

- ▶ Method: Use a variant of Fan et al.'s (2016) "Projected principal component analysis" (PPCA) to estimate a cross section of individual asset returns
- ▶ Details: Suppose return generating processes are relatively stable during a period of time and can be expressed in matrices with time-varying factors and a cross-section of alphas and betas,

$$\mathbf{R} = \boldsymbol{\alpha}\mathbf{1}'_T + \mathbf{B}\mathbf{F}' + \mathbf{E}$$

$$\boldsymbol{\alpha} = \mathbf{G}_\alpha(\mathbf{X}) + \Gamma_\alpha$$

$$\mathbf{B} = \mathbf{G}_\beta(\mathbf{X}) + \Gamma_\beta$$

- ⇒  $\mathbf{F}$ : Factors;  $\mathbf{X}$ : a wide range of firm characteristics (+their polynomials)
- ⇒  $\mathbf{G}_\alpha, \mathbf{G}_\beta$ : a linear function of  $\mathbf{X}$
- ⇒ Estimation: (1) using PPCA to obtain  $\widehat{\mathbf{G}}_\beta$ , (2) estimate  $\widehat{\mathbf{G}}_\alpha$  using regression given  $\widehat{\mathbf{G}}_\beta$ , (3) construct arbitrage portfolios from  $\widehat{\mathbf{G}}_\alpha$
- ▶ Main take-away: There is substantial mispricing. Arbitrage portfolios have significant alphas relative to several popular asset pricing models with annual SR ranging from 0.67 to 1.12

## Comments:

Interesting paper, pushing the big agenda of re-examining the cross section of individual asset returns! Potential major implications on the traditional understanding of risk or anomalies!

## Comments:

Interesting paper, pushing the big agenda of re-examining the cross section of individual asset returns! Potential major implications on the traditional understanding of risk or anomalies!

1. Put the paper in an array of the recent horse race
2. What about the economics?
3. Technical details: Check estimation stability (36 months; cross section subsamples; robustness checks)

## Comment 1: A Horse Race

- ▶ Kelly, Pruitt, and Su, “Characteristics *are* covariances: a unified model of risk and return”, *JFE forthcoming*

## Comment 1: A Horse Race

- ▶ Kelly, Pruitt, and Su, “Characteristics *are* covariances: a unified model of risk and return”, *JFE forthcoming*
  - ⇒ Develop a new method named **Instrumented Principal Component Analysis (IPCA)** which allows for latent factors and time-varying loadings by introducing observable characteristics that instrument for the unobservable dynamic loadings

## Comment 1: A Horse Race

- ▶ Kelly, Pruitt, and Su, “Characteristics *are* covariances: a unified model of risk and return”, *JFE forthcoming*
  - ⇒ Develop a new method named **Instrumented Principal Component Analysis (IPCA)** which allows for latent factors and time-varying loadings by introducing observable characteristics that instrument for the unobservable dynamic loadings
    - Similar: also a wide range of characteristics, also a wide range of individual asset returns instead of portfolio returns, also a variant of PCA, ...

## Comment 1: A Horse Race

- ▶ Kelly, Pruitt, and Su, “Characteristics *are* covariances: a unified model of risk and return”, *JFE forthcoming*
  - ⇒ Develop a new method named **Instrumented Principal Component Analysis (IPCA)** which allows for latent factors and time-varying loadings by introducing observable characteristics that instrument for the unobservable dynamic loadings
    - Similar: also a wide range of characteristics, also a wide range of individual asset returns instead of portfolio returns, also a variant of PCA, ...
  - ⇒ Main take-away: All risk! 5 IPCA factors produce characteristic-associated anomaly intercepts that are small and statistically insignificant.

## Comment 1: A Horse Race

- ▶ Important to discuss why results are very different between the two papers while they share the same primary goal

## Comment 1: A Horse Race

- ▶ Important to discuss why results are very different between the two papers while they share the same primary goal
- ▶ Main differences between the two papers:
  - ⇒ KPS: The dynamics come from the **time series of characteristics** while assets' cross-sectional exposures are **constant**

## Comment 1: A Horse Race

- ▶ Important to discuss why results are very different between the two papers while they share the same primary goal
- ▶ Main differences between the two papers:
  - ⇒ KPS: The dynamics come from the **time series of characteristics** while assets' cross-sectional exposures are **constant**
  - ⇒ This paper: The dynamics come from (1) the changes in the **cross-sectional joint dynamics** between anomalies (alphas), risk-based exposures (betas) and characteristics, as well as (2) the **updating** relation based on a 3-year of monthly rolling windows

## Comment 1: A Horse Race

- ▶ Important to discuss why results are very different between the two papers while they share the same primary goal
- ▶ Main differences between the two papers:
  - ⇒ KPS: The dynamics come from the **time series of characteristics** while assets' cross-sectional exposures are **constant**
  - ⇒ This paper: The dynamics come from (1) the changes in the **cross-sectional joint dynamics** between anomalies (alphas), risk-based exposures (betas) and characteristics, as well as (2) the **updating** relation based on a 3-year of monthly rolling windows
- ▶ But that looks like, results do depend on the choice of methods

## Comment 2: What about the economics?

- ▶ Once advantage of their framework is to generate a time series of **mispricing estimates**,  $G_{\alpha}(\mathbf{X})$ , and **factor exposure estimates**,  $G_{\beta}(\mathbf{X})$

## Comment 2: What about the economics?

- ▶ Once advantage of their framework is to generate a time series of **mispricing estimates**,  $G_{\alpha}(\mathbf{X})$ , and **factor exposure estimates**,  $G_{\beta}(\mathbf{X})$
- ▶ While the paper focuses on establishing the economic magnitude of **the mispricing part** relative to **the systematic risk part** in explaining the cross section of stock returns, I think it is also interesting to show more results on **the systematic part**.

## Comment 2: What about the economics?

- ▶ Once advantage of their framework is to generate a time series of **mispricing estimates**,  $G_\alpha(\mathbf{X})$ , and **factor exposure estimates**,  $G_\beta(\mathbf{X})$
- ▶ While the paper focuses on establishing the economic magnitude of **the mispricing part** relative to **the systematic risk part** in explaining the cross section of stock returns, I think it is also interesting to show more results on **the systematic part**.
- ▶ Why?
  1. Increase the confidence in your estimates
  2. Compare results with KPS
  3. Infer useful information for structural asset pricing and macro theories,  
after carefully controlling the possibility and the degree of mispricing:  
What are the behaviors of the latent PCA factors in the U.S. market?  
Are betas time-varying and what are salient relations between betas and firm characteristics?

# Conclusion

- ▶ I highly recommend it!

# Conclusion

- ▶ I highly recommend it!
- ▶ To make it more convincing:
  1. It might be quite important to formally address some of the competing papers in this recent “horse race”
  2. The model has the power to provide some unique insights into structural asset pricing theory, e.g. systematic risk, time-varying betas, relation between beta and firm characteristics. Some economic discussions would be appreciated
  3. More robustness checks can be done (e.g., rolling window sizes etc.)

Thank You!  
nancy.xu@bc.edu