3/24/2022

The University of Texas at Austin
McCombs School of Business

ML IN FINANCE
CHAPTER 4.4–4.6

Presented by
Batuhan Kuzu
PhD Student in Finance, The University of Texas at Austin

OUT OF SAMPLE WITH FULLY WITHHELD SAMPLE

- Analysis so far: Evaluation of SDFs with OOS data not used in estimation, but penalty parameters picked to maximize OOS performance
- Our anomaly data set likely includes data-mined anomalies and ones that have deteriorated due to learning (McLean and Pontiff 2016)
- Statistical significance of claim that SDF is not characteristics-sparse?
- Conduct estimation of SDF coefficients & penalty choice based on pre-2005 data; use post-2004 for evaluation (2005-2017 OOS period)

OUT OF SAMPLE WITH FULLY WITHHELD SAMPLE

- Constructing SDF is equivalent to finding the MVE portfolio
- Pre-2005 data yield SDF coefficient estimates b_hat = MVE portfolio weights
- Apply b_hat to post 2004 returns: OOS MVE portfolio return
- Test alphas relative to restricted benchmarks
  - Construct MVE portfolio weights from sparse characteristics-based models (e.g., FF 5-factor model) in pre-2005 data and apply weights to post-2004 returns
  - Regress OOS MVE portfolio return on MVE portfolio return of sparse characteristics-based factors to estimate OOS abnormal return

OUT OF SAMPLE WITH FULLY WITHHELD SAMPLE

TABLE 4.1

MVE portfolio's annualized OOS α in the withheld sample (2005–2017)
The table shows annualized alphas (in %) computed from the time-series regression of the SDF-implied OOS-MVE portfolio's returns (based on L^2-shrinkage only) relative to four restricted benchmarks: CAPM, Fama-French six-factor model, optimal sparse model with five factors, and optimal PC-sparse model with five PC-based factors. MVE portfolio returns are normalized to have the same standard deviation as the aggregate market. Standard errors are in parentheses.

SDF factors \ Benchmark	CAPM	FF 6-factor	Char.-sparse	PC-sparse
50 anomaly portfolios	12.35	8.71	9.55	4.60
	(5.26)	(4.94)	(3.95)	(2.22)
80 WFR portfolios	20.05	19.77	17.08	3.63
	(5.26)	(5.29)	(5.05)	(2.93)
1,375 interactions of anomalies	25.00	22.79	21.68	12.41
	(5.26)	(5.18)	(5.03)	(3.26)

EXAMPLES FROM RECENT LITERATURE ON ML METHODS IN CROSS-SECTIONAL AP

	Regularization	Assets	Nonlinearity
SDF models			
Kozak, Nagel, Santosh (2019)	elastic net	char. portfolios PC portfolios	interactions
Kozak (2019)	elastic net	char. portfolios PC portfolios	kernels
Giglio, Feng, and Xiu (2019)	Lasso	char. portfolios	-
DeMiguel et al. (2019)	Lasso	char. portfolios	-
Beta models			
Kelly, Pruitt, Su (2018)	PCA cutoff	indiv. stocks	-
Gu, Kelly and Xiu (2019)	Lasso	char. portfolios	autoencoder neural nets
Return prediction models			
Freyberger, Neuhierl, Weber (2018)	Group lasso	indiv. stocks	splines
Moritz and Zimmerman (2016)	Random forest	indiv. stocks	interactions
Gu, Kelly, Xiu (2018)	many	indiv. stocks	many

EXAMPLES FROM RECENT LITERATURE ON ML METHODS IN CROSS-SECTIONAL AP

	Regularization	Assets	Nonlinearity
SDF models			
Kozak, Nagel, Santosh (2019)	elastic net	char. portfolios PC portfolios	interactions
Kozak (2019)	elastic net	char. portfolios PC portfolios	kernels
Giglio, Feng, and Xiu (2019)	Lasso	char. portfolios	-
DeMiguel et al. (2019)	Lasso	char. portfolios	-
Beta models			
Kelly, Pruitt, Su (2018)	PCA cutoff	indiv. stocks	-
Gu, Kelly and Xiu (2019)	Lasso	char. portfolios	autoencoder neural nets
Return prediction models			
Freyberger, Neuhierl, Weber (2018)	Group lasso	indiv. stocks	splines
Moritz and Zimmerman (2016)	Random forest	indiv. stocks	interactions
Gu, Kelly, Xiu (2018)	many	indiv. stocks	many

ML IN CROSS-SECTIONAL AP: SUMMARY

- ML well-suited to address challenges in cross-sectional AP
- Given the existing factor zoo, there is little point in analyzing a few return predictors in isolation: ML methods here to stay
- Outcomes much noisier than typical ML application: bringing in informed priors (economic restrictions & motivations) important
- Sparsity depends on rotation: can be quite sparse with PCs, but less so with original characteristics-based factors

3/24/2022

The University of Texas at Austin
McCombs School of Business

100

FACTORS THAT FIT THE TIME SERIES AND CROSS-SECTION OF STOCK RETURNS

Martin Lettau & Markus Pelger (RFS 2020)

Presented by
Batuhan Kuzu
PhD Student in Finance, The University of Texas at Austin

MOTIVATION

- Fundamental question of asset pricing:

  What are risk factors and how are they priced?

- Current state of literature:

  “Factor zoo” with over 300 potential asset pricing factors published.

- Open questions:

  - Which factors are really important in explaining expected returns?
  - Which are subsumed by others?

- How do we determine important factors?

GOALS OF THE PAPER

- Bring order into this ‘factor chaos’
- Summarize the pricing information of a larger number of assets with a small number of factors
- ‘Let he data speak’ rather than sorting stocks according to pre-defined characteristics
- Identify factors that
  1) explain time-series variation
  2) explain the cross-section of risk premia
  3) have high Sharpe ratios

INTUITION

- Arbitrage Pricing Theory (APT): Prices of financial assets should be explained by systematic risk factors
- The APT links the times series and cross-section of returns but is silent about the factors
- Factors are identified by either
  - time series methods (Principal Components)
  - cross-sectional methods (Fama-French models)
- **This paper combines *time-series* and *cross-sectional* objectives**
- Risk factors should
  1) explain time-series variation
  2) explain cross-section of risk premia
  3) have high Sharpe ratios

CONTRIBUTION OF THE PAPER

- This paper: Estimation approach for finding risk factors
- Key elements of estimator:

  1) Statistical factors instead of pre-specified (and potentially miss-specified) factors

  2) Uses information from large panel data sets: Many assets with many time observations

  3) Searches for factors explaining asset prices (explain differences in expected returns) not only co-movement in the data

  4) Allow time-variation in factor structure

CONTRIBUTION OF THE PAPER

- Asymptotic distribution theory for weak and strong factors
  => No ‘blackbox approach’
- Estimator discovers ‘weak’ factors with high Sharpe ratios
  => high Sharpe ratio factors important for asset pricing and investment
- Estimator strongly dominates conventional approach (Principal Component Analysis (PCA))
  - PCA does not find all high Sharpe ratio factors
- Empirical results:
  - New factors much smaller pricing errors in- and out- of sample than benchmark (PCA, 5 Fama-French factors etc.)
  - 3 times higher Sharpe ratio then benchmark factors (PCA)

A FACTOR MODEL OF ASSET RETURNS

- Observe excess returns $X_{t,i}$ of N assets over T time periods:

$$X_{t,i} = \underbrace{\underset{1\times K}{F_t}^{\top}}_{factors} \underbrace{\underset{K\times 1}{\Lambda_i}}_{loadings} + \underbrace{e_{t,i}}_{idiosyncratic} \qquad i = 1, \dots, N \quad t = 1, \dots, T$$

$$\underbrace{X}_{T\times N} = \underbrace{F}_{T\times K} \underbrace{\Lambda^{\top}}_{K\times N} + \underbrace{e}_{T\times N}$$

- T time-series observation (large)
- N test assets (large)
- K systematic factors (fixed)
- F, Λ and e are unknown

A STATISTICAL MODEL OF ASSET RETURNS: SYSTEMATIC FACTORS

Systematic and non-systematic risk (F and e uncorrelated):

$$\mathrm{Var}(X) = \underbrace{\Lambda\,\mathrm{Var}(F)\,\Lambda^\top}_{systematic} + \underbrace{\mathrm{Var}(e)}_{non-systematic}$$

$\Rightarrow$ Systematic factors should explain a large portion of the variance

$\Rightarrow$ Idiosyncratic risk can be weakly correlated

Estimation via PCA: Minimize the unexplained variance:

$$\min_{\Lambda,F} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{ti} - F_t\,\Lambda_i^\top)^2$$

A STATISTICAL MODEL OF ASSET RETURNS: SYSTEMATIC FACTORS

Arbitrage-Pricing Theory (APT): The expected excess return is explained by the risk-premium of the factors:

$$\mathrm{E}[X_i] = \mathrm{E}[F]\,\Lambda_i^\top$$

$\Rightarrow$ Systematic factors should explain the cross-section of expected returns

Estimation: Minimize cross-sectional pricing error

$$\min_{\Lambda,F} \frac{1}{N} \sum_{i=1}^{N} \left(\bar{X}_i - \bar{F}\,\Lambda_i^\top\right)^2$$

$$\bar{X}_i = \frac{1}{T} \sum_{t=1}^{T} X_{t,i}$$

$$\bar{F} = \frac{1}{T} \sum_{t=1}^{T} F_t$$

RISK-PREMIA PCA (RP-PCA) ESTIMATOR

RP-PCA : Minimize jointly the unexplained variance and pricing error

$$\min_{\Lambda,F} \quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{ti} - F_t\Lambda_i^\top)^2 + \gamma \frac{1}{N}\sum_{i=1}^{N}\left(\bar{X}_i - \bar{F}\Lambda_i^\top\right)^2$$

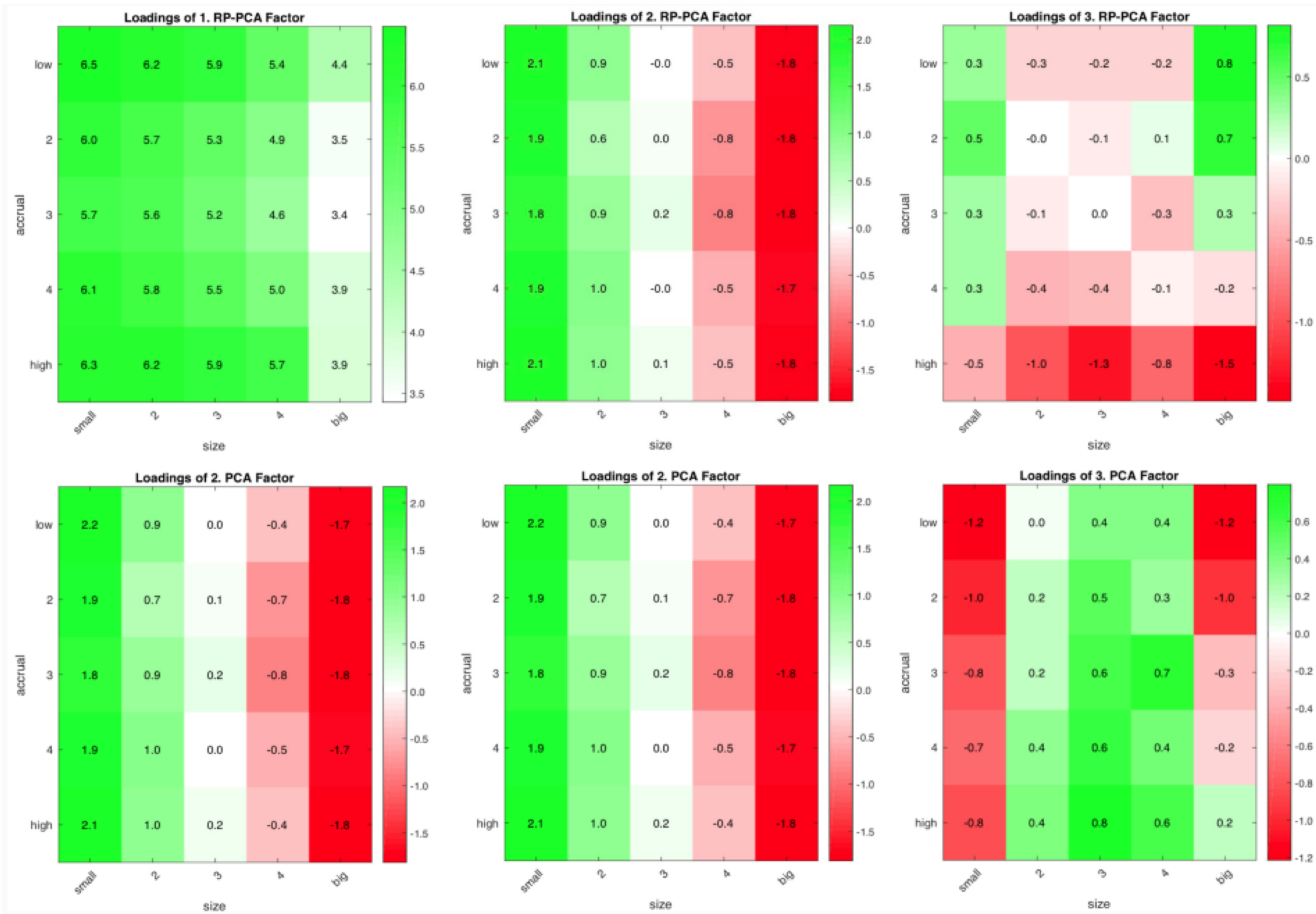
1. Combine variation and pricing error criterion functions:
   - Select factors with small cross-sectional pricing errors (alpha's).
   - Protects against spurious factor with vanishing loadings as it requires the time-series errors to be small as well.
2. Penalized PCA: Search for factors explaining the time-series but penalizes low Sharpe-ratios.
3. Special cases:
   - $\gamma = -1$: PCA with demeaned returns and factors
   - $\gamma = 0$: PCA, returns and factors not demeaned

ILLUSTRATION: SIZE AND ACCRUAL PORTFOLIOS

	In-sample			Out-of-sample		
	SR	RMS α	Idio.Var.	SR	RMS α	Idio.Var.
RP-PCA	0.24	0.12	6.11	0.21	0.11	6.75
PCA	0.13	0.14	5.92	0.11	0.14	6.72
FF-long/sort	0.21	0.12	7.90	0.11	0.12	7.11

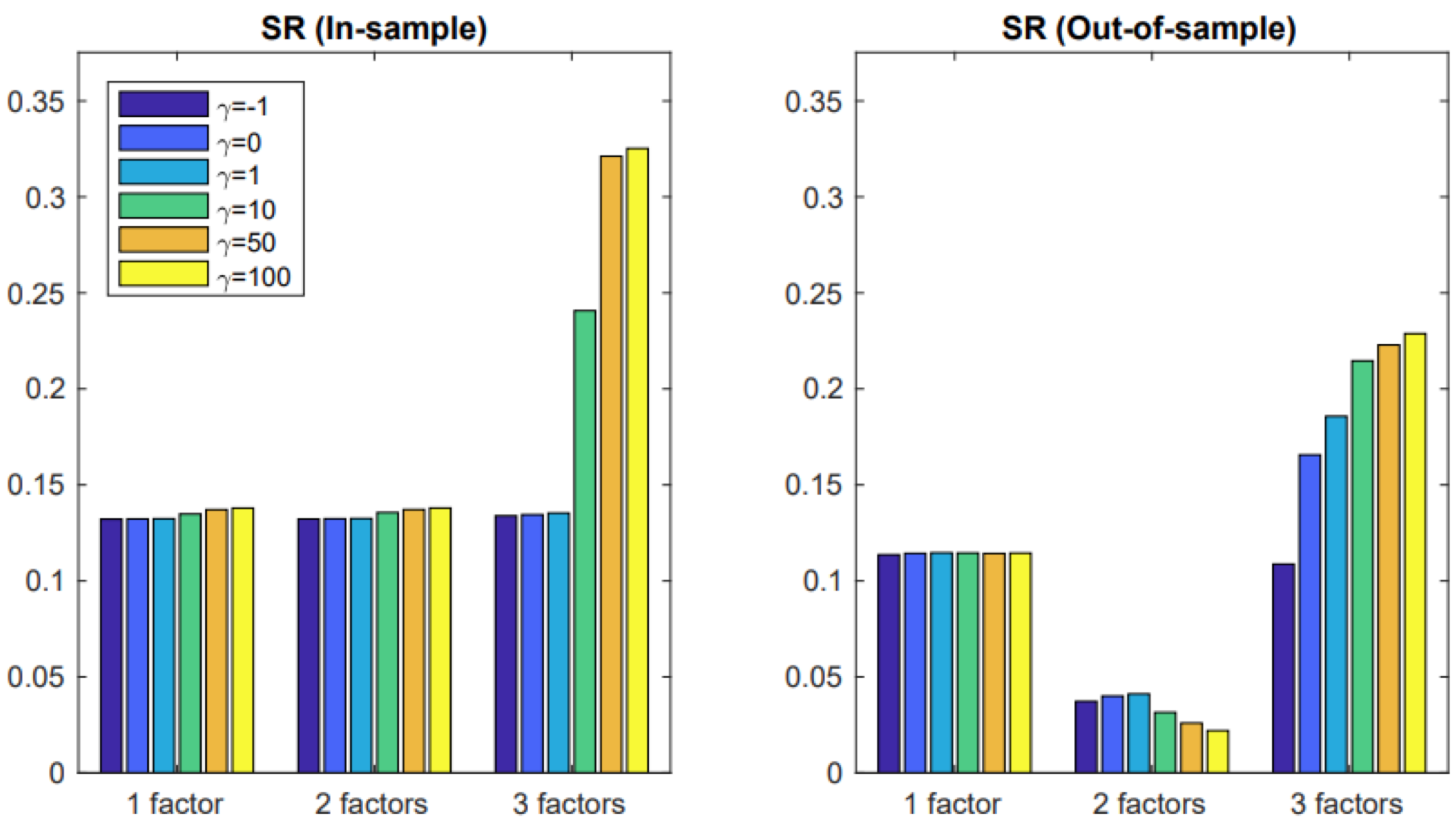
- Pricing error $\alpha_i = E[X_i] - E[F]\Lambda_i^\top$
- RMS α: Root-mean-squared pricing errors $\sqrt{\frac{1}{N}\sum_{i=1}^{N}\alpha_i{}^2}$
- Idiosyncratic Variation: $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{t,i} - F_t^\top\Lambda_i)^2$

LOADINGS FOR STATISTICAL FACTORS: SIZE AND ACCRUAL PORTFOLIOS



RP-PCA detects accrual factor while 3rd PCA factor is noise.

MAXIMAL SHARPE RATIO (SIZE AND ACCRUAL)



⇒ 1st and 2nd PCA and RP-PCA factors the same.

⇒ RP-PCA detects 3rd factor (accrual) for $\gamma > 10$.

SOME THEORY: WEAK VS STRONG FACTORS

Recall the factor model:

$$X_{t,i} = F_t^\top \Lambda_i + e_{t,i}$$

- Strong factors ($\frac{1}{N}\Lambda^\top\Lambda$ bounded)
  - Strong factors affect most assets (proportional to N), e.g. market factor
  - Strong factors lead to exploding eigenvalues

  $\Rightarrow$ RP-PCA always more efficient than PCA
- Weak factors ($\Lambda^\top\Lambda$ bounded)
  - Weak factors affect a smaller fraction of assets, e.g. value factor
  - Weak factors lead to large but bounded eigenvalues

  $\Rightarrow$ RP-PCA detects weak factors which cannot be detected by PCA

STATISTICAL THEORY: STRONG FACTORS

- PCA under assumptions of Bai (2003): (up to rotation)
  - Asymptotically $\hat{\Lambda}$ behaves like OLS regression of F on X.
  - Asymptotically $\hat{F}$ behaves like OLS regression of Λ on $X^\top$.
- RP-PCA under slightly stronger assumptions as in Bai (2003):
  - Asymptotically $\hat{\Lambda}$ behaves like OLS regression of FW on XW with $W = \left(I_T + \gamma \frac{\mathbb{1}\mathbb{1}^\top}{T}\right)$ and $\mathbb{1}$ is a $T \times 1$ vector of 1's .
  - Asymptotically $\hat{F}$ behaves like OLS regression of Λ on X.
- Asymptotic Efficiency: Choose RP-weight γ to obtain smallest asymptotic variance of estimators
  - RP-PCA and PCA are both consistent
  - RP-PCA (i.e. $\gamma > -1$) always more efficient than PCA

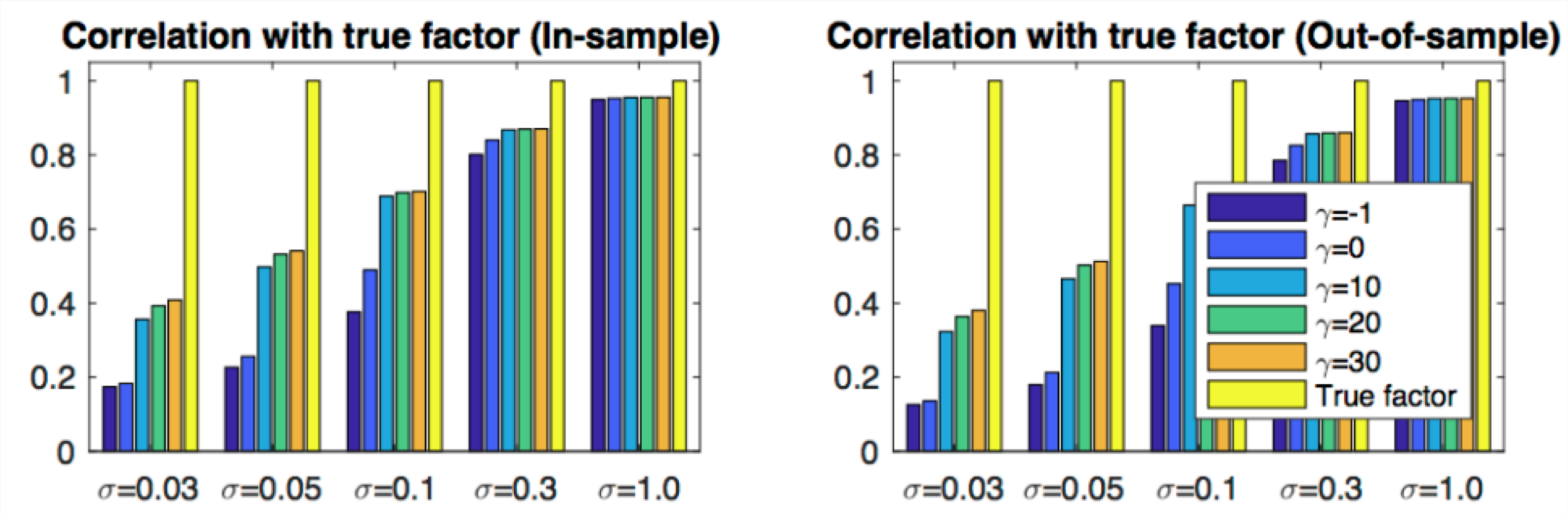
STATISTICAL THEORY: WEAK FACTORS

- Weak factors either have a small variance or affect a smaller fraction of assets
- $\Lambda^\top \Lambda$ bounded (after normalizing factor variances)
- Statistical methods: Spiked covariance models, random matrix theory
- Main results:
  - Weak factors are much more difficult to detect than strong factors
  - Neither PCA nor RP-PCA yield consistent estimators of true factors
  - But $\widehat{Corr}(F, \hat{F})$ is maximized for $\gamma > -1$
  - RP-PCA strictly dominates PCA
  - ... especially when a weak factor has a high SR
- The number of factors can be determined by the spectrum of eigenvalues

SIMULATION

- Fix number of assets (N) and sample size (T)
- Presentation: Single weak factor ($K = 1$) with $SR = 0.8$
- Vary volatility of factor (the higher the volatility, the stronger the factor)
- Vary weight γ
- Compare
    - SR of estimated factor (IS and OOS)
    - RMS of XS pricing errors
    - correlation with true factor

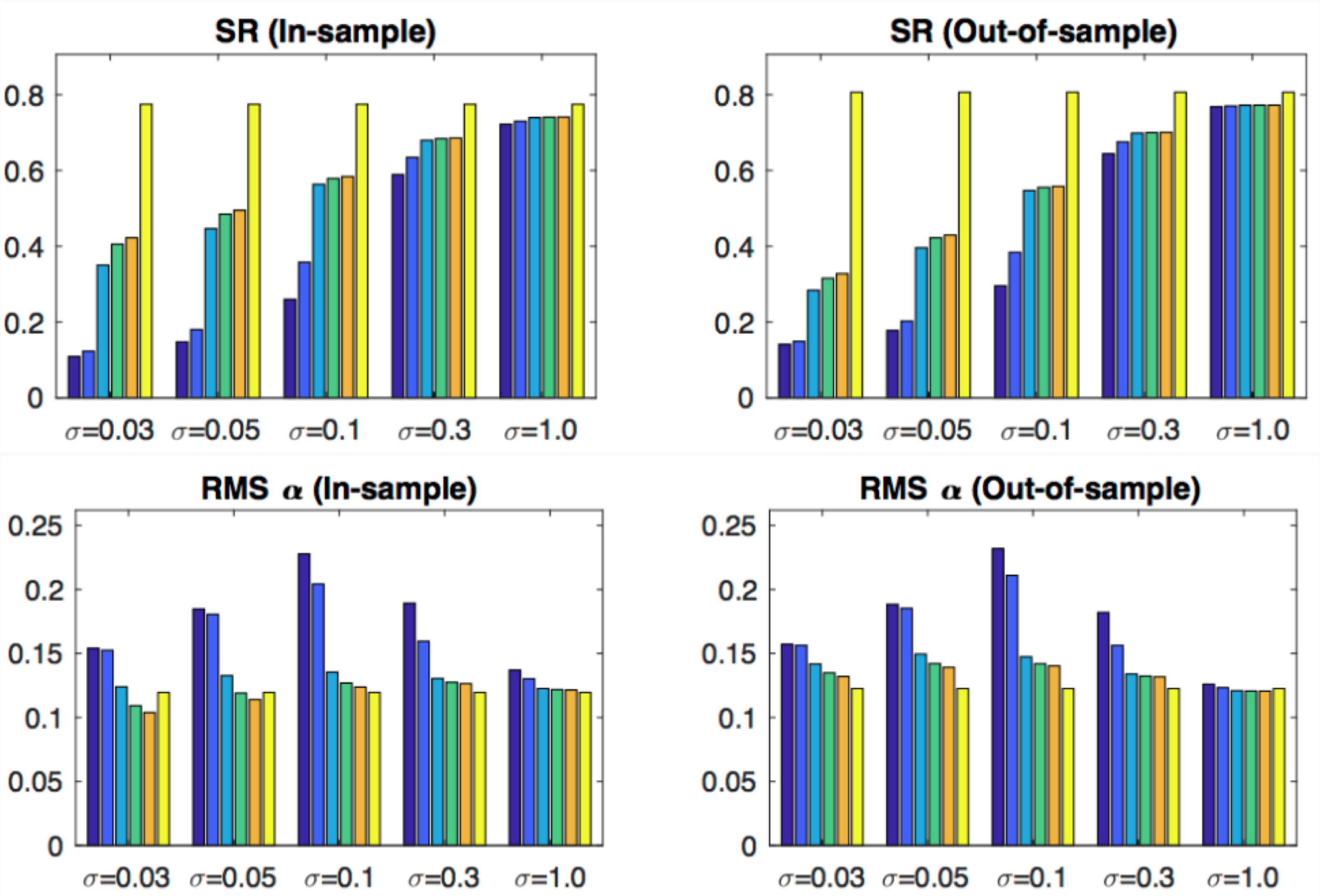
SIMULATION


Correlation with true factor (In-sample)
1
0.8
0.6
0.4
0.2
0
σ=0.03
σ=0.05
σ=0.1
σ=0.3
σ=1.0
Correlation with true factor (Out-of-sample)
γ=-1
γ=0
γ=10
γ=20
γ=30
True factor

The University of Texas at Austin
McCombs School of Business

100

SIMULATION



EMPIRICAL RESULTS: DOUBLE-SORTED PORTFOLIOS

- Data
  - Monthly return data from 07/1963 to 05/2017 ($T = 647$)
  - 13 sets of double-sorted portfolios (each consisting of 25 portfolios)
- Factors
  1. **PCA:** $K = 3$
  2. **RP-PCA:** $K = 3$ and $\gamma = 100$
  3. **FF-Long/Short** factors: market + two specific anomaly long-short factors

SHARPE RATIOS AND PRICING ERRORS (IN-SAMPLE)

	Sharpe-Ratio			α		
	RPCA	PCA	FF-L/S	RPCA	PCA	FF-L/S
Size and BM	0.25	0.22	0.21	0.13	0.13	0.14
BM and Investment	**0.24**	**0.17**	**0.24**	**0.08**	**0.11**	**0.12**
BM and Profits	0.23	0.20	0.24	0.10	0.12	0.16
Size and Accrual	**0.33**	**0.13**	**0.21**	**0.06**	**0.14**	**0.12**
Size and Beta	0.25	0.24	0.23	0.06	0.07	0.10
Size and Investment	0.32	0.26	0.21	0.10	0.11	0.22
Size and Profits	0.22	0.21	0.25	0.06	0.06	0.16
Size and Momentum	0.25	0.19	0.18	0.15	0.16	0.17
Size and ST-Reversal	0.28	0.25	0.24	0.16	0.17	0.35
Size and Idio. Vol.	0.35	0.31	0.32	0.15	0.16	0.16
Size and Total Vol.	0.34	0.30	0.31	0.16	0.16	0.16
Profits and Invest.	0.28	0.24	0.30	0.10	0.12	0.11
Size and LT-Reversal	0.22	0.18	0.16	0.11	0.13	0.16

SHARPE RATIOS AND PRICING ERRORS (OUT-OF-SAMPLE)

	Sharpe-Ratio			α		
	RPCA	PCA	FF-L/S	RPCA	PCA	FF-L/S
Size and BM	0.23	0.18	0.16	0.18	0.19	0.19
BM and Investment	**0.21**	**0.15**	**0.24**	**0.13**	**0.17**	**0.17**
BM and Profits	0.19	0.17	0.23	0.16	0.17	0.19
Size and Accrual	**0.24**	**0.09**	**0.11**	**0.08**	**0.14**	**0.12**
Size and Beta	0.22	0.20	0.16	0.08	0.09	0.09
Size and Investment	0.30	0.23	0.17	0.13	0.14	0.16
Size and Profits	0.22	0.20	0.20	0.10	0.1	0.17
Size and Momentum	0.18	0.12	0.08	0.17	0.18	0.19
Size and ST-Reversal	0.22	0.17	0.23	0.22	0.23	0.25
Size and Idio. Vol.	0.37	0.30	0.28	0.17	0.18	0.18
Size and Total Vol.	0.35	0.28	0.27	0.18	0.20	0.19
Profits and Invest.	0.31	0.25	0.29	0.12	0.15	0.14
Size and LT-Reversal	0.16	0.10	0.04	0.13	0.14	0.14

SINGLE SORTED PORTFOLIOS

Portfolio Data

- Monthly return data from 07/1963 to 12/2016 ($T = 638$) for $N = 370$ portfolios
- Kozak, Nagel and Santosh (2017) data: 370 decile portfolios sorted according to 37 anomalies
- Factors:
  1. **RP-PCA**: $K = 6$ and $\gamma = 100$.
  2. **PCA**: $K = 6$
  3. **Fama-French 5:** The five factor model of Fama-French (market, size, value, investment and operating profitability, all from Kenneth French's website).
  4. **Proxy factors**: RP-PCA and PCA factors approximated with 5% of largest position.

SINGLE SORTED PORTFOLIOS

	In-sample			Out-of-sample		
	SR	RMS α	Idio. Var.	SR	RMS α	Idio. Var.
RP-PCA	**0.66**	**0.15**	**2.73**	**0.53**	**0.11**	**3.19**
PCA	0.28	0.15	2.70	0.22	0.14	3.19
Fama-French 5	0.32	0.23	4.97	0.31	0.21	4.62

- RP-PCA strongly dominates PCA and Fama-French 5 factors
- Results hold out-of-sample.

SINGLE SORTED PORTFOLIOS: MAXIMAL SHARPE RATIO

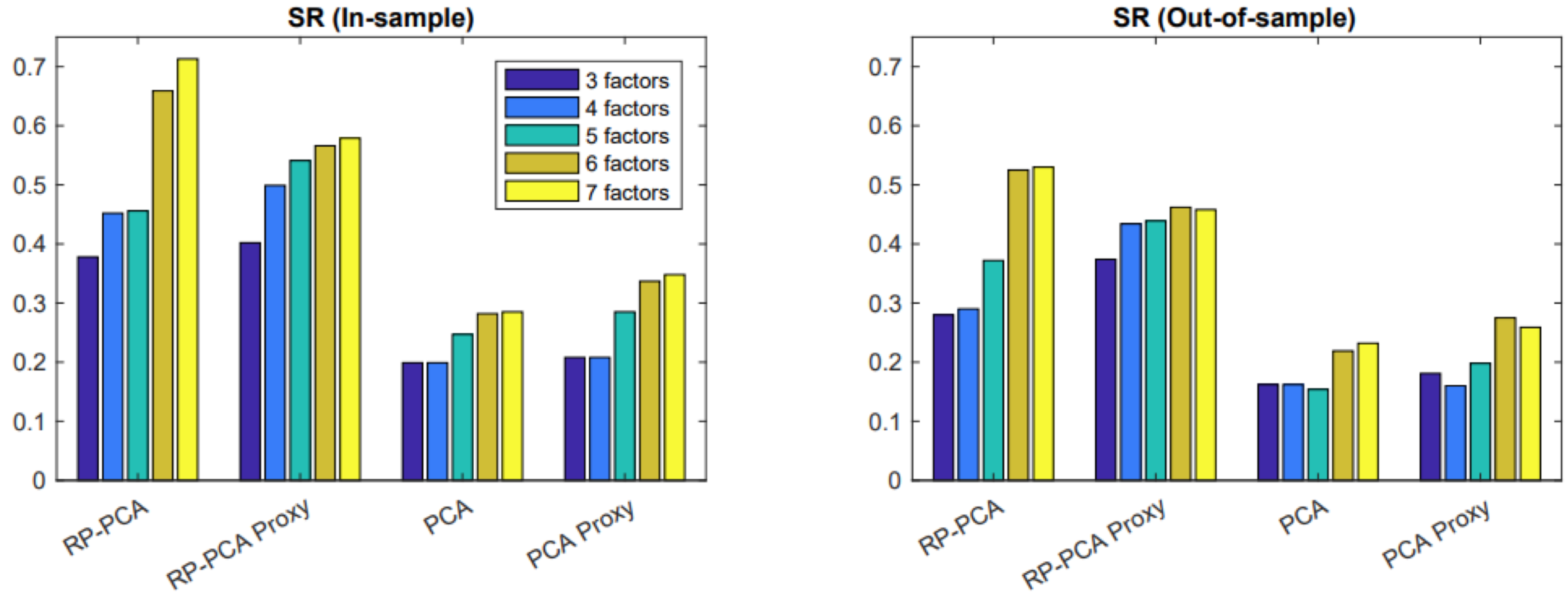


Figure 1: Maximal Sharpe-ratios.

Spike in Sharpe ratio for 6 factors

SINGLE SORTED PORTFOLIOS: PRICING ERROR

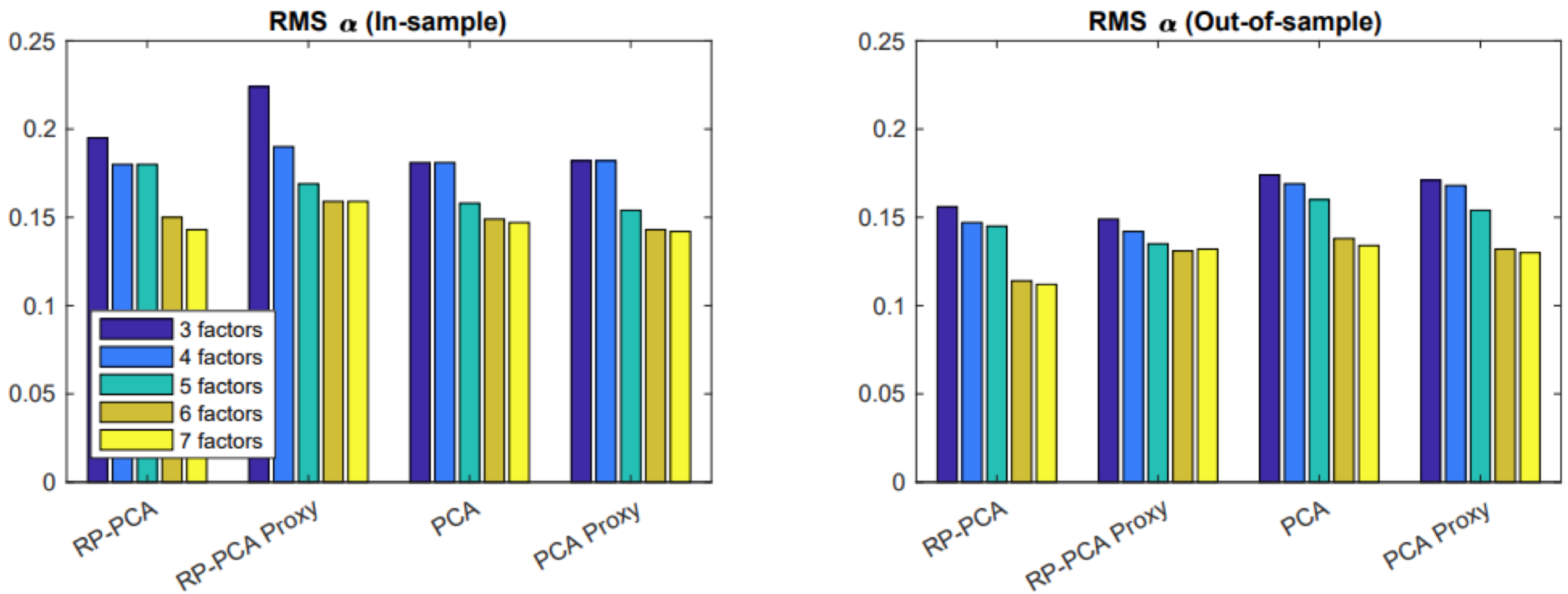


Figure 2: Root-mean-squared pricing errors.

RP-PCA has smaller out-of-sample pricing errors

SINGLE SORTED PORTFOLIOS: IDIOSYNCRATIC VARIATION

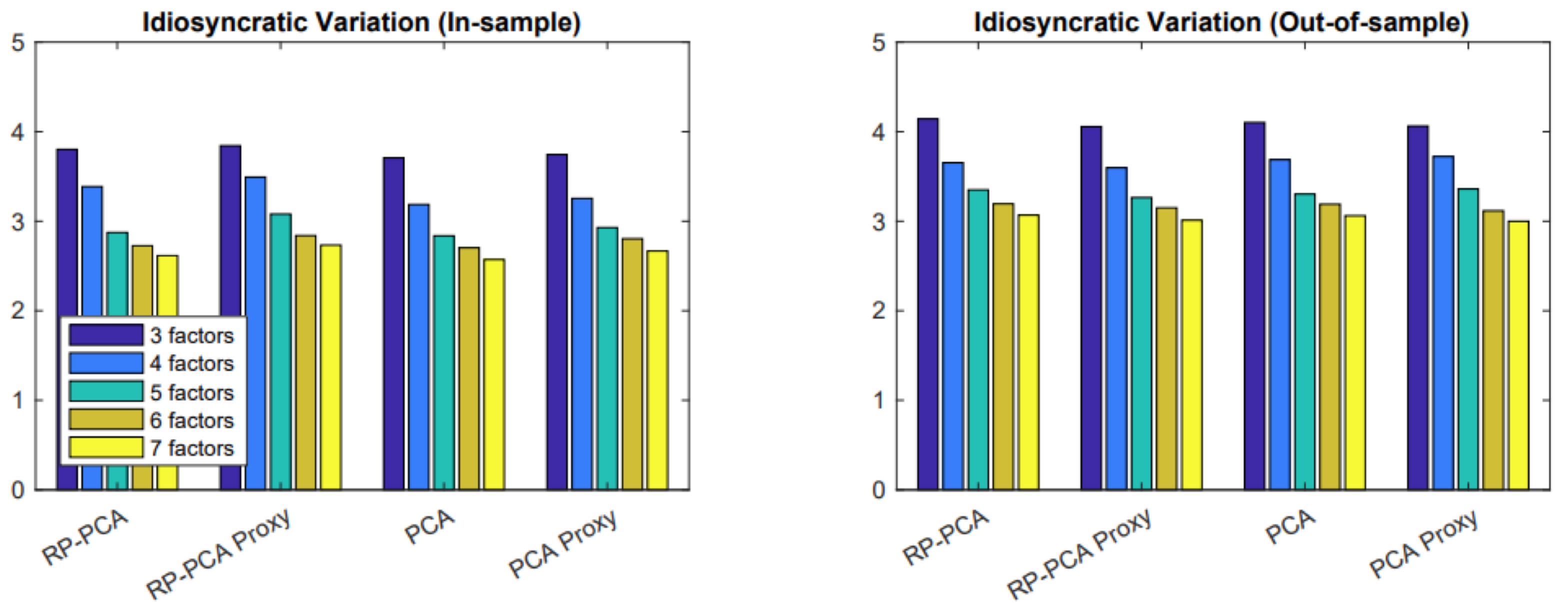


Figure 3: Unexplained idiosyncratic variation.

Unexplained variation similar for RP-PCA and PCA

TIME STABILITY: GENERALIZED CORRELATIONS

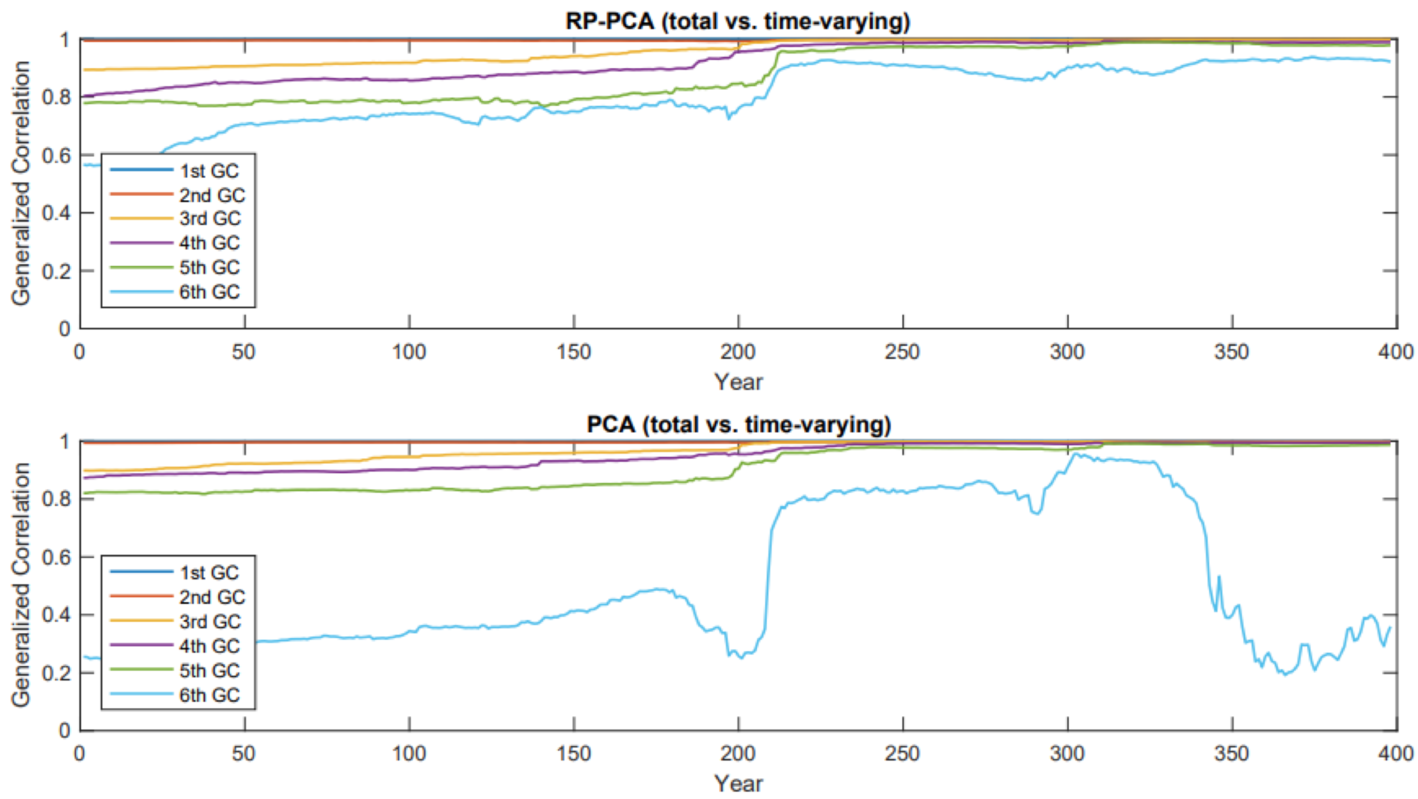


Figure 4: Generalized correlations between loadings estimated on the whole time horizon $T = 638$ and a rolling window with 240 months for $K = 6$ factors.

CONCLUSION

Methodology

- Estimator for estimating priced latent factors from large data sets
- Combines variation and pricing criterion function
- Asymptotic theory under weak and strong factor model assumption
- Detects weak factors with high Sharpe-ratio
- More efficient than conventional PCA

Empirical Results

- Strongly dominates PCA of the covariance matrix.
- Potential to provide benchmark factors for horse races.

THANK YOU FOR LISTENING!