Factors that Fit the Time Series and Cross-Section of Stock Returns

Martin Lettau [1] Markus Pelger [2]

[1]UC Berkeley

[2]Stanford University

November 30th 2018

NBER Asset Pricing Meeting

Motivation

- Fundamental question: What are risk factors and how are they priced?
- Current state of the literature: "Factor zoo" with 300+ potential asset pricing factors!
- Goal of this paper: Bring order into "factor chaos"
  - ⇒ Summarize the pricing information of a large number of assets with a small number of factors
- Conventional methods (Principal Component Analysis PCA): Only time-series co-movement, ignores risk premia
- Ross' Arbitrage Pricing Theory (APT) links time series and cross section: Risk premia depend on exposure to non-diversifiable factors
- Our method: PCA + APT = Risk Premium PCA (RP-PCA)
- RP-PCA estimates factors that
  1. explain the time-series variation
  2. explain the cross-section of risk premia
  3. have high Sharpe-ratios

A factor model of asset returns

- Observe excess returns X_{nt} of N assets over T time periods:

$$X_{nt} = \mathbf{F}_t \mathbf{B}_n^\top + e_{nt} \qquad n = 1, ..., N \quad t = 1, ..., T$$

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

- T: time-series observation (large)
- N: test assets (large)
- K: systematic factors (fixed)
- $\mathbf{F}$, $\mathbf{B}$ and $\mathbf{e}$ are unknown

A statistical model of asset returns: Systematic factors

- Systematic and non-systematic risk:

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

  - Systematic factors explain large portion of variance
  - Idiosyncratic risk only weakly correlated
- Estimation via standard PCA: Minimize the unexplained variance

$$\min_{\mathbf{B},\mathbf{F}} \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} (X_{nt} - \mathbf{F}_t \mathbf{B}_n^\top)^2$$

- Easy to estimate: Eigenvectors/values of covariance matrix of $\mathbf{X}$
  - PCA factors capture common co-movements
  - ... but usually do not fit cross-sectional mean returns and low SR
  - Example: Industry factors

A statistical model of asset returns: Risk premia

- Arbitrage-Pricing Theory (APT, Ross 1976):
  The expected excess return is explained by the risk-premium of the factors:

$$\mathrm{E}[\mathbf{X}_n] = \mathbf{B}_n \,\mathrm{E}[\mathbf{F}]$$

  ⇒ Systematic factors should explain the cross-section of expected returns

- Estimation: Minimize cross-sectional pricing error

$$\min_{\mathbf{B},\mathbf{F}} \frac{1}{N} \sum_{n=1}^{N} \left(\bar{\mathbf{X}}_n - \bar{\mathbf{F}}\, \mathbf{B}_i^\top \right)^2$$

with $\bar{\mathbf{X}}_n = \frac{1}{T} \sum_{t=1}^{T} X_{nt}$

Risk-Premium PCA (RP-PCA) estimator

Risk-Premium PCA: PCA + APT=RP-PCA:

$$\min_{\mathbf{B},\mathbf{F}} \underbrace{\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}(X_{nt} - \mathbf{F}_t \mathbf{B}_n^\top)^2}_{\text{unexplained TS variation}} + \gamma \underbrace{\frac{1}{N}\sum_{n=1}^{N}\left(\bar{\mathbf{X}}_n - \bar{\mathbf{F}}\mathbf{B}_n^\top\right)^2}_{\text{XS pricing error}}$$

- γ is the weight on the APT mean restriction
- Estimation: Apply standard PCA to the matrix with overweighted mean

$$\frac{1}{T}\mathbf{X}^\top\mathbf{X} + \gamma\bar{\mathbf{X}}\bar{\mathbf{X}}^\top.$$

- Special case: $\gamma = -1$: (Standard) PCA on covariance matrix

Interpretation of RP-PCA

1. Combines variation and pricing error criterion functions:
   - 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 (consistent with APT)
3. Information interpretation: (GMM interpretation)
   - PCA of a covariance matrix uses only the second moment.
   - RP-PCA combines first and second moments efficiently.
4. Signal-strengthening: Intuitively the matrix $\frac{1}{T}X^\top X + \gamma \bar{X}\bar{X}^\top$ converges to

$$\mathbf{B}\left(\text{Var}(\mathbf{F}) + (1+\gamma)\text{E}[\mathbf{F}]\text{E}[\mathbf{F}]^\top\right)\mathbf{B}^\top + \text{Var}(e)$$

The signal of weak factors with a small variance can be "pushed up" by their mean with the right γ.

Theoretical results

- Types of factors:
  - Strong factors affect all assets: MKT
  - Weak factors affect either some assets “strongly or all asset “weakly”: Long-short factors
- Theory: Companion paper (Lettau and Pelger (2018))
  - Strong factors: RP-PCA more efficient
    - RP-PCA with $\gamma = 0$ is optimal but PCA works as well
  - Weak factors: RP-PCA dominates PCA
    - RP-PCA is able to detect factors that are missed by PCA
    - Optimal $\gamma \approx 10 - 20$
    - Novel results based on random matrix theory for non-zero means
- Paper:
  - Criterion to select number of factors
  - Extensive simulation study, in-sample and out-of-sample

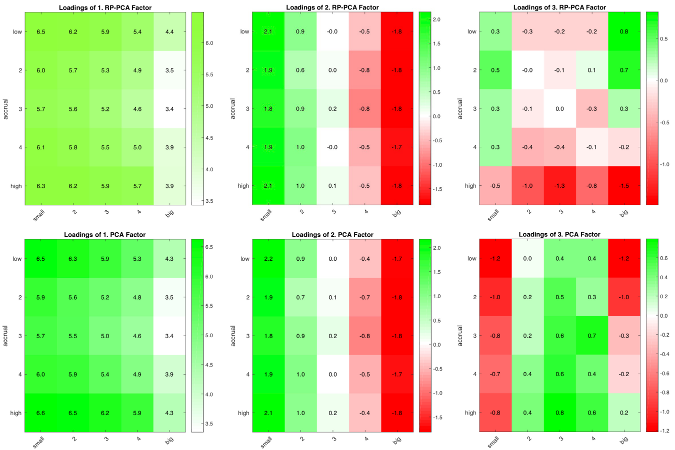
Illustration (Size and accrual)

Model	Out-of-sample		
	SR	RMS α	Idio.Var.
RP-PCA	0.21	0.11	6.75
PCA	0.11	0.14	6.72
FF-long/sort	0.11	0.12	7.11

Table: $K = 3$ factors and $\gamma = 10$.

- Monthly returns of 25 double-sorted portfolios
- SR $=$ Sharpe-ratio of mean-variance efficient portfolio
- $RMS\alpha$ $=$ Root Mean Squared XS pricing error
- RP-PCA factors have higher SRs than PCA factors and FF factors
- RP-PCA model has lower pricing errors than PCA and FF factors
- RP-PCA has $< 1\%$ higher idiosyncratic TS variance than PCA

Loadings for statistical factors: Size and accrual portfolios



RP-PCA:	$SR_1 = .11$	$SR_2 = .03$	$SR_3 = .23$
PCA:	$SR_1 = .10$	$SR_2 = .02$	$SR_3 = .11$

Large Cross-section of Single-Sorted Portfolios

- Kozak, Nagel and Santosh (2017) data: 370 decile portfolios sorted according to 37 anomalies
- Monthly return data from 07/1963 to 12/2017 ($T = 650$) for $N = 370$ portfolios
- Two cases:
  - $N = 74$ extreme 1/10 decile portfolios
  - All $N = 370$ portfolios
- Factors:
  1. RP-PCA: $K = 5$ and $\gamma = 10$.
  2. PCA: $K = 5$
  3. Fama-French 5

Single-sorted portfolios

Model	Out-of-sample		
	SR	RMS α	Idio.Var.
RP-PCA	0.45	0.12	12.70%
PCA	0.17	0.14	12.56%
Fama-French 5	0.31	0.21	13.66%

Table: Maximal Sharpe-ratios, root-mean-squared pricing errors and unexplained idiosyncratic variation. $K = 5$ factors and $\gamma = 10$.

- RP-PCA strongly dominates PCA and Fama-French 5 factors
- RP-PCA has larger SR and smaller pricing errors
- RP-PCA explains (almost) the same variation as PCA
- Results hold out-of-sample.

Maximal Sharpe-ratio

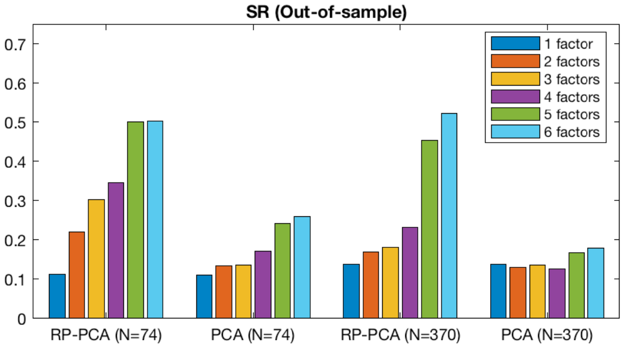


Figure: Maximal Sharpe-ratios for extreme ($N = 74$) and all ($N = 370$) deciles.

- RP-PCA: 5th factor (green) has big effect on SR
- PCA: higher order factors contribute less to SR
- Extreme deciles capture most of the pricing information

Choice of γ: Maximal Sharpe-ratio

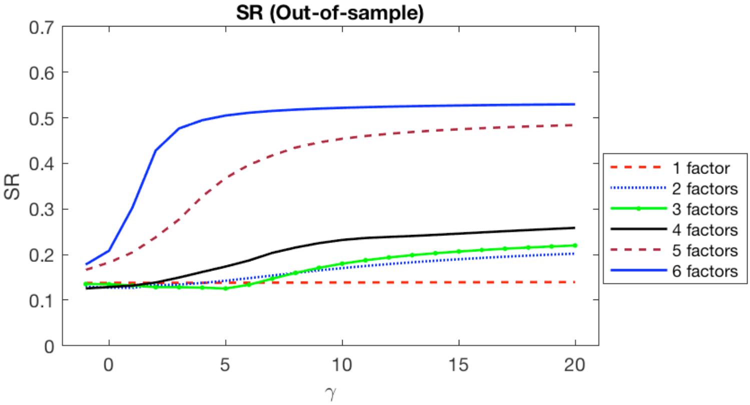


Figure: Maximal Sharpe-ratios for different γ and K for $N = 370$.

$\Rightarrow$ Strong increase in Sharpe-ratios for $\gamma \geq 10$ and $K = 5$.

Unexplained Variation

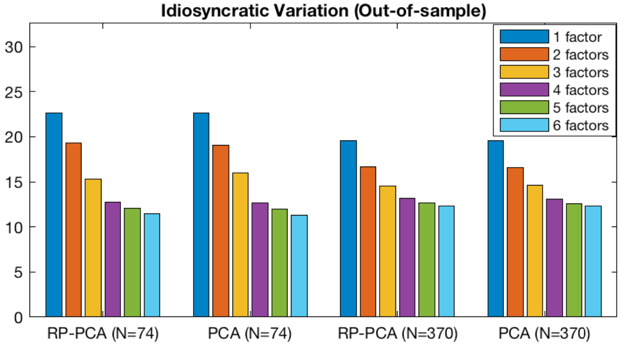
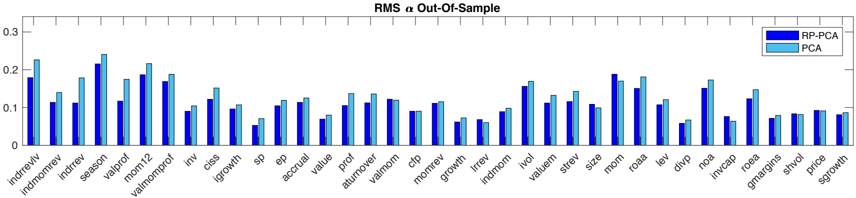


Figure: Idiosyncratic variation for extreme ($N = 74$) and all ($N = 370$) deciles.

- RP-PCA captures (almost) as much common variation as PCA factors

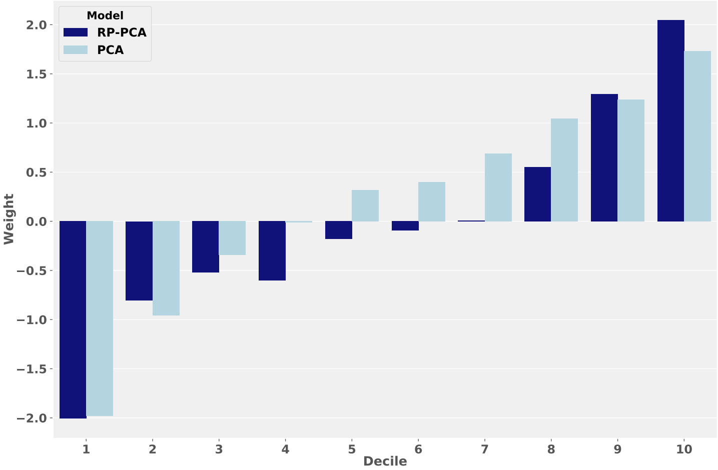
RMS of Pricing Errors α's: $N = 370$



- Characteristics sorted according to Sharpe-ratios of LS portfolios

⇒ High SR characteristics significantly better priced by RP-PCA

Composition of factors

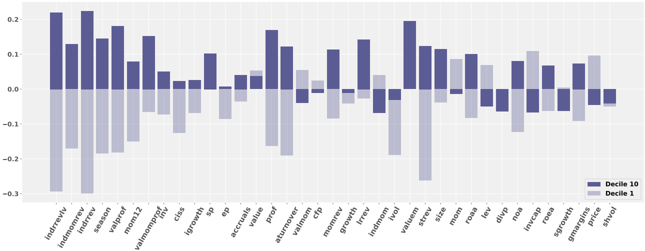


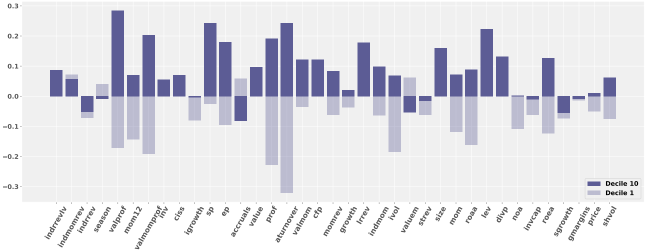
- Loading weights within deciles for all characteristics.

⇒ Almost all weights on extreme deciles.

Stochastic Discount Factors for $N = 74$

Order portfolios by SR! (top RP-PCA, bottom PCA)





RP-PCA vs. PCA Factors

	RP-PCA			PCA		
	Mean	Variance	SR	Mean	Variance	SR
1. Factor	11.67	7295.35	0.14	11.56	7387.22	0.13
2. Factor	2.65	222.62	0.18	1.66	241.03	0.11
3. Factor	0.46	213.34	0.03	0.23	207.49	0.02
4. Factor	2.40	125.92	0.21	1.52	132.57	0.13
5. Factor	2.76	39.10	0.44	0.78	49.30	0.11

- Normalize loadings $\mathbf{B}^\top \mathbf{B} = I_K$.
- Fifth RP-PCA factor weak and high SR

Interpretation of factors

	RP-PCA	PCA
1. Factor	long-only market proxy	long-only market proxy
2. Factor	value/growth + interactions	value/growth
3. Factor	momentum+profitability	?
4. Factor	momentum+momentum interaction	momentum
5. Factor	high SR long/short portfolios	?

Note: Factors are comprised mostly of "classic" anomaly portfolios

Generalized Correlation with Ad-Hoc Long/Short Portfolios

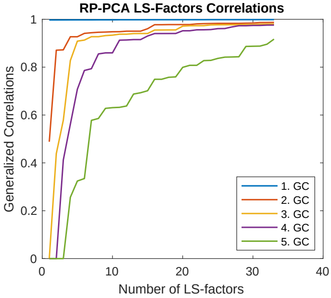


Figure: Generalized correlations of statistical factors with increasing number of long- short anomaly factors.

- First LS-factor is the market factor and LS-factors added incrementally based on the largest accumulative absolute loading.

⇒ Ad-hoc long-short factors do not span statistical factors.

Time-stability of loadings

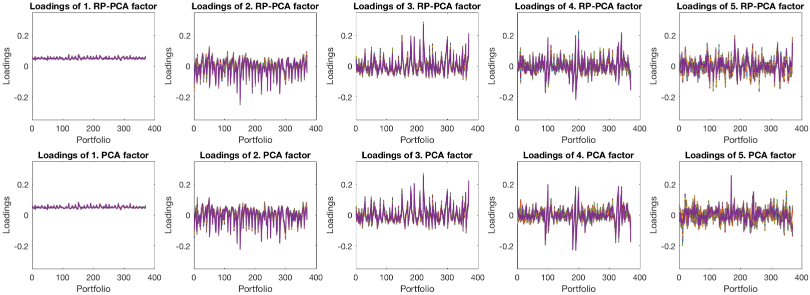


Figure: Time-varying rotated loadings for the first six factors. Loadings are estimated on a rolling window with 240 months.

⇒ RP-PCA stable over time

Next steps

- So far: Pre-sorted portfolios, constant factor-weights and loadings
- Goal: Construct factors using individual stocks without any pre-sorts
- Obstacle: Factor loadings of individual stocks are likely to vary over time
- Example: Momentum exposure
- Current work: Time-varying loadings
- General idea behind all time-varying factor models
  1. Create portfolios = projection based on characteristics
  2. Estimate constant loading model on projected data
  3. Invert model to obtain loadings that vary with characteristics
- Our approach
  - More general portfolios (projection) based on decision trees
  - RP-PCA to extract factors on tree portfolios
  - Invert model with second dimension reduction

Conclusion

Methodology

- PCA-based estimator with cross-sectional mean information
- Easy to implement: Conventional PCA on 2nd-moment matrix plus mean term
- RP-PCA can detect weak factors that are missed by standard PCA
- More efficient than conventional PCA

Empirical Results

- RP-PCA dominates standard PCA
- RP-PCA higher-order factors with high SRs
- RP-PCA loadings put more weight on high-SR characteristics
- Robust in out-of-sample estimation
- 5 economically meaningful asset pricing factors
- Redundancy in characteristic information

Factors 1: Long-only ("Mkt")

Factor 1 (sorted by Category)

	value	valuem	divp	ep	cfp	sp	valmom	valmomprof	valprof	mom	mom12	indmom	lrrev	strev	momrev	indmomrev	indrrev	indrrevlv	inv
RP-PCA10	5.0	5.7	3.5	4.9	5.0	5.6	4.8	5.4	5.3	5.4	5.8	4.7	5.8	6.3	6.0	5.0	7.1	5.6	5.2
PCA10	4.6	5.5	3.1	4.5	4.7	5.2	4.4	4.7	4.8	5.1	5.3	4.4	5.6	6.5	5.8	4.1	6.8	4.8	4.9
RP-PCA1	4.5	4.6	5.0	5.6	4.8	4.4	4.9	5.0	4.6	6.0	5.5	5.1	5.2	4.4	5.2	3.5	4.6	3.4	5.1
PCA1	4.6	4.6	5.2	6.1	5.0	4.6	5.2	5.4	5.0	6.2	6.6	5.4	5.4	4.8	5.5	4.0	5.2	4.0	5.4

	invcap	igrowth	growth	sgrowth	prof	roaa	roea	noa	gmargins	aturnover	size	ivol	accruals	ciss	lev	price	season	shvol
RP-PCA10	4.1	5.5	5.0	4.7	4.4	4.3	4.4	5.1	4.2	4.7	5.4	3.1	5.4	4.0	5.1	3.9	5.9	3.1
PCA10	3.9	5.5	4.9	4.8	4.1	4.2	4.3	5.1	4.2	4.5	5.3	2.9	5.4	3.6	5.0	3.9	5.5	2.9
RP-PCA1	5.7	5.4	5.4	5.5	3.7	5.6	6.0	4.7	4.9	3.6	3.9	6.5	5.3	5.0	4.5	6.6	4.6	6.5
PCA1	6.0	5.8	5.7	5.7	3.8	6.1	6.4	5.1	5.0	3.8	3.9	7.6	5.7	5.4	4.6	7.0	5.0	7.0

- Factor 1: Long in (almost) all portfolios

Factor 2: Value and value-interaction

Factor 2 (sorted by Category)

	value	valuem	divp	ep	cfp	sp	valmom	valmomprof	valprof	mom	mom12	indmom	lrrev	strev	momrev	indmomrev	indrrev	indrrevlv	inv
RP-PCA10	1.3	0.5	1.5	1.2	1.1	1.0	1.2	1.0	1.0	0.3	0.3	0.6	0.6	-0.8	0.2	1.3	-0.3	1.1	0.4
PCA10	2.2	2.9	2.3	1.5	1.9	2.0	0.8	0.1	1.4	-1.0	-1.1	-0.5	2.1	0.7	1.3	-0.1	0.6	0.5	0.6
RP-PCA1	-1.2	-1.0	-1.0	-1.5	-1.3	-1.2	-1.4	-1.3	-1.1	-1.2	-2.3	-1.2	-1.2	-0.9	-1.2	-0.8	-1.4	-0.8	-1.1
PCA1	-1.7	-1.8	-1.0	-0.4	-1.6	-1.5	-0.3	0.2	-0.5	1.2	1.9	0.9	-1.4	-0.3	-1.1	0.5	-0.2	-0.3	-0.6

	invcap	igrowth	growth	sgrowth	prof	roaa	roea	noa	gmargins	aturnover	size	ivol	accruals	ciss	lev	price	season	shvol
RP-PCA10	0.7	-0.3	0.2	-0.1	-0.2	-0.7	-0.5	-0.8	-0.8	0.0	0.1	0.4	-0.7	0.9	0.8	-0.4	-0.1	0.4
PCA10	1.0	0.3	0.9	0.9	-1.2	-1.3	-1.0	-1.1	-1.4	-0.2	1.5	-0.1	-0.6	0.7	1.7	-1.1	-0.5	0.4
RP-PCA1	-1.7	-1.5	-1.5	-1.2	-0.1	-1.2	-0.9	-1.2	-0.2	-0.5	-0.3	-2.5	-1.4	-0.9	-1.2	-1.1	-0.8	-1.7
PCA1	-1.6	-0.7	-1.0	-1.1	0.4	0.3	0.7	-0.1	0.5	0.1	-0.6	0.8	-0.5	0.1	-1.7	2.5	0.5	-0.7

- RP-PCA: Long/short in value and value-interaction portfolios
- PCA: Mostly value portfolios

Factor 3: Momentum

- RP-PCA: Momentum-related portfolios
- PCA: No clear pattern

Factor 4: Momentum-Interaction

Factor 4 (sorted by Category)

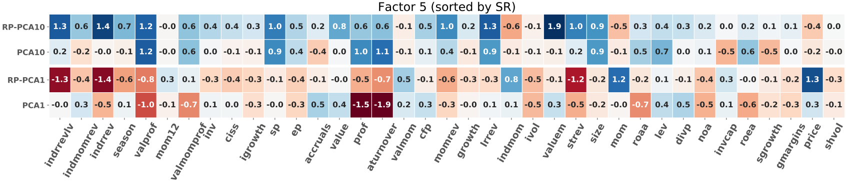
	value	valuem	divp	ep	cfp	sp	valmom	valmomprof	valprof	mom	mom12	indmom	lrrev	strev	momrev	indmomrev	indrrev	indrrevlv	inv
RP-PCA10	0.2	-0.7	-1.1	-0.3	-0.1	0.3	1.4	1.9	0.6	1.8	1.8	1.6	1.2	-0.2	0.8	-0.2	0.6	-0.6	0.4
PCA10	0.5	-1.7	-0.0	0.4	0.4	0.3	2.2	2.1	0.5	2.4	2.1	2.4	0.5	-1.1	0.2	0.7	-0.5	-0.3	0.5
RP-PCA1	-0.6	-0.1	-0.5	1.0	-0.4	-0.5	-1.4	-1.4	-0.6	-1.4	-1.2	-1.5	-0.2	0.8	0.5	-1.6	1.0	-0.6	-0.1
PCA1	-0.6	-0.0	-0.5	0.3	-0.5	-0.5	-2.1	-1.9	-0.4	-2.7	-2.7	-2.4	-0.2	0.9	0.4	-1.2	0.9	0.3	-0.1

	invcap	igrowth	growth	sgrowth	prof	roaa	roea	noa	gmargins	aturnover	size	ivol	accruals	ciss	lev	price	season	shvol
RP-PCA10	-0.5	0.9	0.4	0.1	-0.5	-0.5	-0.7	0.2	-0.5	-0.4	2.1	-1.2	0.5	-0.7	-1.2	-0.4	0.3	-1.1
PCA10	0.2	0.6	0.4	0.1	-0.2	-0.4	-0.4	-0.0	-0.4	-0.2	1.2	-0.3	0.1	-0.0	-0.5	0.2	0.1	-0.4
RP-PCA1	0.8	0.6	0.2	0.6	-0.6	1.1	1.5	-0.2	0.8	-0.6	-1.0	2.1	0.4	-0.4	-0.4	1.7	0.4	1.7
PCA1	-0.2	0.1	-0.2	0.1	-0.1	0.3	0.7	-0.4	0.8	-0.4	-0.5	0.3	-0.1	-0.3	-0.5	-0.5	0.4	0.7

- RP-PCA and PCA: Momentum and momentum-interaction portfolios

Factor 5: High SR

Note: Order portfolios by SR instead of categories!



RP-PCA: Long in highest SR portfolios

PCA: Asset Turnover and Profitability

Individual stocks

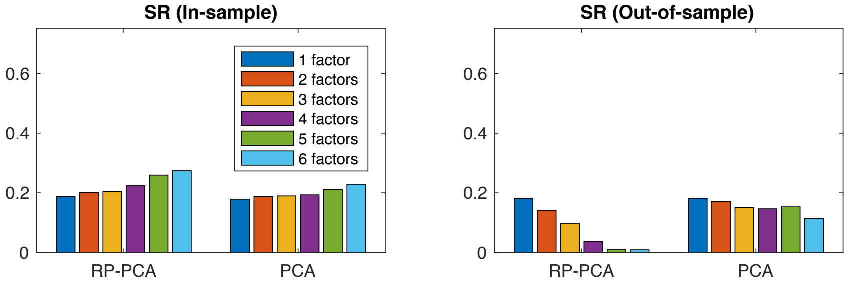


Figure: Stock price data ($N = 270$ and $T = 500$): Maximal Sharpe-ratios for different number of factors. RP-weight $\gamma = 10$.

- Stock price data from 01/1972 to 12/2016 ($N = 270$ and $T = 500$)

$\Rightarrow$ Out-of-sample performance collapses

$\Rightarrow$ Constant loading model inappropriate

Time-stability of loadings of individual stocks

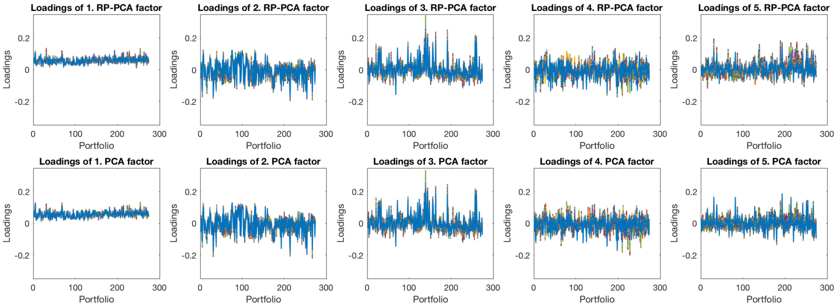


Figure: Stock price data: Generalized correlations between loadings estimated on the whole time horizon and a rolling window

Time-stability of loadings of individual stocks

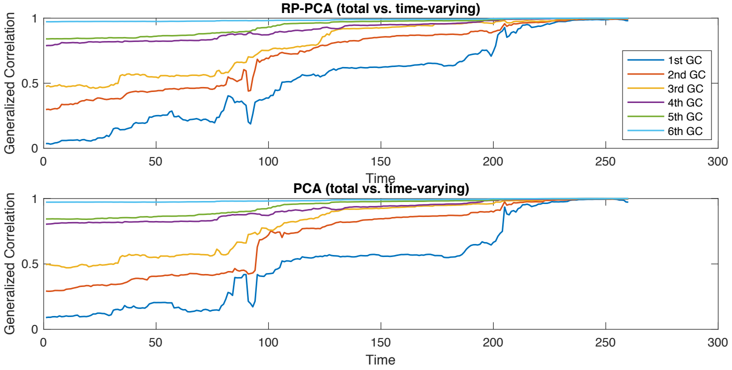


Figure: Stock price data ($N = 270$ and $T = 500$): Generalized correlations between loadings estimated on the whole time horizon and a rolling window with 240 months.

Portfolio data: Average Pricing Errors

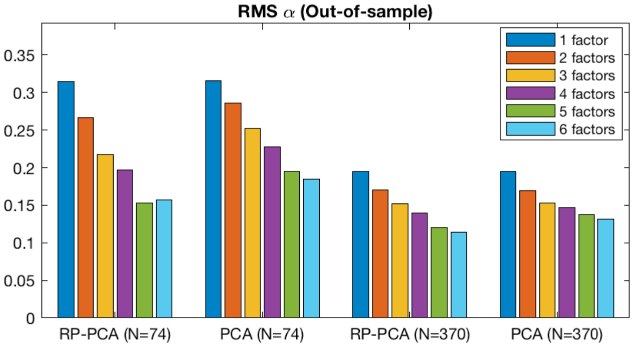


Figure: Maximal Sharpe-ratios for extreme ($N = 74$) and all ($N = 370$) deciles.

- RP-PCA has smaller pricing errors.

Single-sorted portfolios

	In-sample			Out-of-sample		
	SR	RMS α	Idio. Var.	SR	RMS α	Idio. Var.
RP-PCA	0.53	0.14	10.76%	0.45	0.12	12.70%
PCA	0.24	0.14	10.66%	0.17	0.14	12.56%
Fama-French 5	0.32	0.23	13.56%	0.31	0.21	13.66%

Table: Maximal Sharpe-ratios, root-mean-squared pricing errors and unexplained idiosyncratic variation. $K = 5$ factors and $\gamma = 10$.

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

Portfolio data: Average Pricing Errors

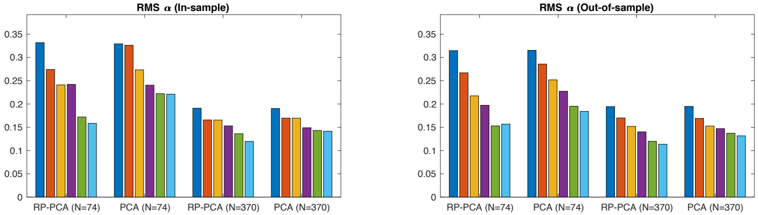
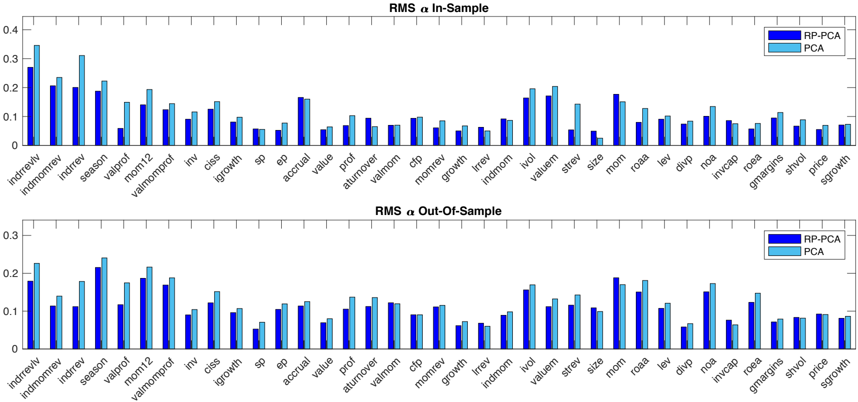


Figure: Maximal Sharpe-ratios for extreme ($N = 74$) and all ($N = 370$) deciles.

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

RMS of TS α's: $N = 74$



All 370 portfolios: RP-PCA

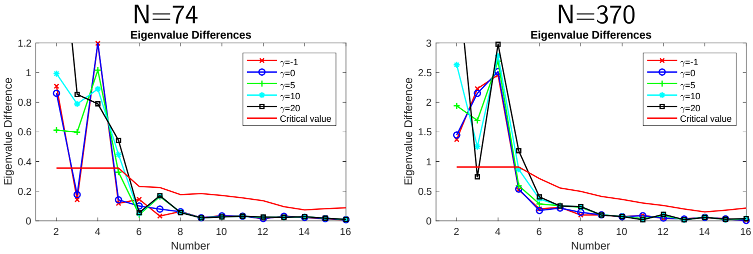
	indrrevlv	indmomrev	indrrev	season	valprof	mom12	valmomprof	inv	ciss	igrowth	sp	ep	accruals	value	prof	aturnover	valmom	cfp	momrev	growth	lrrev	indmom	ivol	valuem	strev	size	mom	roaa	lev	divp	noa	invcap	roea	sgrowth	gmargins	price	shvol
10	1.7	1.2	1.3	1.0	1.5	0.9	1.4	0.3	0.3	0.1	0.8	0.3	0.1	0.4	1.6	1.1	-0.0	0.0	0.8	-0.0	1.1	-0.0	0.1	0.9	0.5	0.7	0.3	0.9	0.2	-0.0	0.7	-0.4	0.7	0.6	-0.0	-0.1	-0.6
9	1.5	0.7	1.4	0.4	1.1	0.0	0.9	0.3	0.2	0.2	0.5	0.2	0.6	-0.1	1.1	1.0	-0.5	0.2	0.3	0.2	0.4	0.2	0.1	0.7	1.0	0.5	-0.2	0.2	-0.0	-0.2	0.3	-0.3	0.5	-0.1	0.1	-0.1	-0.1
8	1.0	0.4	0.9	0.1	1.0	0.1	0.4	0.2	-0.0	0.1	0.5	0.2	-0.1	-0.3	0.7	0.1	-0.5	-0.2	0.4	-0.1	0.4	-0.2	0.1	-0.2	0.8	0.5	-0.4	0.0	-0.2	-0.3	0.3	-0.1	0.2	0.1	0.0	-0.1	-0.1
7	0.5	0.2	0.5	0.1	0.7	-0.2	0.2	0.2	-0.0	0.3	-0.0	0.0	-0.0	-0.3	0.0	0.2	-0.3	-0.6	0.1	-0.1	0.1	0.0	-0.0	-0.2	0.4	0.4	-0.2	-0.5	-0.3	0.0	0.1	-0.3	-0.3	-0.1	0.1	-0.2	-0.3
6	0.1	0.2	0.2	-0.1	0.7	-0.2	-0.3	-0.0	0.2	0.0	-0.1	-0.2	-0.1	-0.6	-0.2	0.3	0.0	-0.3	0.4	-0.1	0.1	0.6	0.1	-0.7	0.2	0.3	0.0	-0.5	-0.6	0.1	0.0	-0.1	-0.5	-0.5	-0.2	0.2	0.3
5	-0.1	0.1	-0.1	0.0	0.7	-0.1	0.0	0.2	0.3	-0.0	-0.1	-0.3	0.1	-0.6	-0.5	0.3	0.2	-0.2	-0.0	-0.1	-0.0	0.2	0.2	-0.5	0.0	0.2	0.1	-0.5	-0.7	-0.2	0.1	0.0	-0.6	-0.3	0.1	0.2	0.5
4	-0.2	-0.3	-0.4	-0.4	-0.0	-0.0	-0.0	-0.1	-0.0	-0.2	-0.1	-0.1	0.1	-0.8	-1.0	0.2	0.3	-0.1	-0.2	0.1	-0.0	-0.1	-0.1	-0.7	-0.3	0.1	0.3	-0.5	-0.2	-0.2	-0.7	0.3	-0.5	-0.4	-0.2	0.5	0.1
3	-0.6	-0.1	-0.8	-0.3	-0.7	0.0	-0.0	0.2	-0.1	0.1	-0.4	0.0	0.1	0.0	-1.2	-0.2	0.3	-0.2	-0.0	0.3	-0.1	0.2	0.2	-0.4	-0.3	-0.0	0.4	-0.5	-0.5	0.1	-0.0	-0.0	-0.6	-0.2	-0.3	0.3	0.3
2	-0.8	-0.3	-1.1	-0.6	-0.8	-0.4	0.1	-0.2	-0.3	0.3	-0.2	-0.3	0.0	0.3	-1.2	-0.8	0.5	0.5	-0.2	0.4	-0.5	0.1	-0.3	0.0	-1.0	-0.2	0.5	-0.4	0.1	-0.1	-0.4	0.2	-0.6	-0.5	-0.1	0.3	0.0
1	-1.7	-1.1	-1.9	-1.1	-1.4	-1.7	-0.7	-0.7	-0.8	-0.6	0.2	-0.7	-0.3	0.5	-0.9	-1.2	0.2	0.3	-0.7	-0.4	-0.2	-0.1	-1.7	0.5	-1.5	-0.2	0.2	-0.6	0.6	0.0	-1.1	0.7	-0.6	-0.6	-0.4	-0.1	-0.0

All 370 portfolios: PCA

	indrrevlv	indmomrev	indrrev	season	valprof	mom12	valmomprof	inv	ciss	igrowth	sp	ep	accruals	value	prof	aturnover	valmom	cfp	momrev	growth	lrrev	indmom	ivol	valuem	strev	size	mom	roaa	lev	divp	noa	invcap	roea	sgrowth	gmargins	price	shvol
10	0.5	0.5	-0.5	-0.2	1.4	0.9	1.4	0.3	0.2	-0.1	1.2	1.3	-0.7	0.7	0.8	0.9	1.2	0.7	0.4	0.3	1.0	1.1	0.4	-1.0	-0.3	1.1	1.0	-0.0	1.9	1.1	-0.3	0.1	0.3	-0.5	0.4	-0.1	-0.3
9	0.5	0.8	-0.0	-0.0	0.9	0.8	1.0	0.4	0.1	-0.2	1.3	0.9	-0.8	0.4	0.4	1.0	0.7	0.4	0.4	0.1	0.6	1.0	0.2	-0.6	-0.0	1.0	0.6	-0.2	0.9	0.5	-0.7	0.1	0.1	-0.6	0.3	-0.0	0.0
8	0.2	0.7	-0.1	-0.1	0.8	0.7	0.9	0.2	0.5	0.1	1.1	0.7	-0.7	0.3	-0.2	0.0	0.4	0.4	0.4	0.1	0.7	1.0	0.4	-0.3	0.2	0.9	0.7	-0.1	0.5	0.3	0.1	-0.0	0.3	-0.6	0.1	0.1	0.1
7	0.5	0.4	0.1	0.3	0.7	0.4	0.5	-0.4	0.2	-0.3	0.8	0.4	-0.2	0.1	-0.6	0.2	0.1	0.2	0.3	-0.3	0.6	0.7	0.4	-0.5	0.2	0.9	0.6	-0.1	0.2	0.1	-0.1	0.1	0.2	-0.1	0.4	-0.0	0.0
6	0.2	0.5	0.1	0.2	0.4	0.3	0.3	-0.1	0.0	-0.4	0.8	0.4	-0.4	-0.2	-0.4	-0.1	-0.0	0.4	0.2	-0.1	0.5	0.6	0.5	-0.5	0.2	0.7	0.2	-0.3	0.2	0.3	-0.4	-0.1	-0.1	-0.4	0.4	0.1	-0.0
5	0.5	0.4	-0.0	0.2	0.2	0.2	-0.0	-0.4	-0.1	-0.2	0.5	-0.0	0.0	-0.4	-0.5	0.3	0.1	0.2	0.4	0.1	0.3	0.4	0.1	-0.6	0.2	0.5	-0.1	-0.2	-0.1	0.6	0.1	0.3	0.1	-0.2	0.4	0.0	-0.0
4	0.4	0.2	0.1	0.3	0.0	-0.2	-0.4	-0.4	-0.0	0.0	0.3	0.1	0.4	-0.2	-0.8	-0.3	0.1	0.4	0.2	0.0	0.2	-0.0	-0.2	-0.4	0.1	0.4	-0.2	0.0	-0.2	0.4	-0.4	0.2	-0.4	0.1	0.3	-0.3	-0.1
3	0.4	0.1	-0.0	0.4	-0.9	-0.3	-0.6	-0.6	-0.1	-0.2	0.1	0.0	0.1	0.0	-0.9	-0.8	0.0	0.1	0.2	-0.2	0.2	0.0	-0.2	-0.3	-0.0	0.1	-0.3	0.7	-0.2	0.3	-0.3	-0.3	-0.4	0.0	0.2	-0.1	0.0
2	0.1	-0.1	-0.1	0.2	-1.2	-1.0	-0.9	-0.6	0.0	-0.2	-0.2	-1.0	0.1	-0.1	-0.5	-1.2	-0.3	0.3	-0.1	-0.4	-0.0	-0.5	-0.4	-0.1	-0.1	0.1	-0.9	0.8	-0.0	0.1	-0.4	-0.4	-0.3	0.5	0.2	-0.2	-0.7
1	0.5	-0.1	-0.1	0.2	-1.0	-1.7	-1.7	-0.3	0.2	-0.7	-0.6	-0.5	-0.2	-0.3	-0.7	-1.4	-1.2	-0.2	-0.5	-0.6	-0.5	-1.3	-0.8	0.0	-0.0	-0.3	-1.6	-0.5	-0.3	0.0	-0.8	-0.9	-0.4	0.2	-0.3	-0.8	-0.7

Number of factors

Onatski (2010): Eigenvalue-ratio test



- RP-PCA: 5 factors
- PCA: 4 factors

Time-stability: Generalized correlations

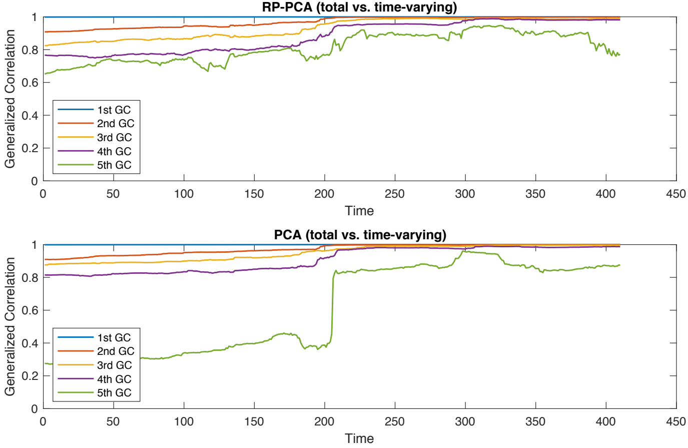


Figure: Generalized correlations between loadings estimated on the whole time horizon $T = 650$ and a rolling window with 240.

Signal of factors: Existence of weak factors

	PCA	RP-PCA ($\gamma = 10$)	FF5
σ_1^2	8.05	8.05	8.00
σ_2^2	0.27	0.27	0.21
σ_3^2	0.21	0.21	0.17
σ_4^2	0.14	0.14	0.03
σ_5^2	0.05	0.05	0.02
σ_6^2	0.03	0.03	0.00

Table: Variance signal for different factors

- Largest eigenvalues of $\frac{1}{N}\Lambda\Sigma_F\Lambda^\top$ normalized by the average idiosyncratic variance $\sigma_e^2 = \frac{1}{N}\sum_{i=1}^N \sigma_{e,i}^2$

$\Rightarrow$ Higher factors are weak.

Signal of factors: Existence of weak factors

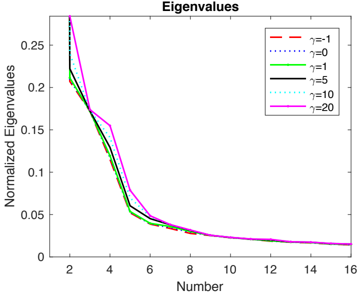


Figure: Largest eigenvalues of the matrix $\frac{1}{N}\left(\frac{1}{T}X^\top X + \gamma \bar{X}\bar{X}^\top\right)$.

- LEFT: Eigenvalues are normalized by division through the average idiosyncratic variance $\sigma_e^2 = \frac{1}{N}\sum_{i=1}^N \sigma_{e,i}^2$.
- RIGHT: Eigenvalues are normalized by the corresponding PCA ($\gamma = -1$) eigenvalues.

$\Rightarrow$ Higher factors have weak variance but high mean signal.

Interpreting factors: Generalized correlations with proxies

	RP-PCA	PCA
1. Gen. Corr.	1.00	1.00
2. Gen. Corr.	0.99	0.99
3. Gen. Corr.	0.98	0.99
4. Gen. Corr.	0.94	0.94
5. Gen. Corr.	0.77	0.89

Table: Generalized correlations of statistical factors with proxy factors (portfolios of 5% of assets).

- Problem in interpreting factors: Factors only identified up to invertible linear transformations.
- Generalized correlations close to 1 measure of how many factors two sets have in common.

⇒ Proxy factors approximate statistical factors.

Single-sorted portfolios: Maximal Sharpe-ratio

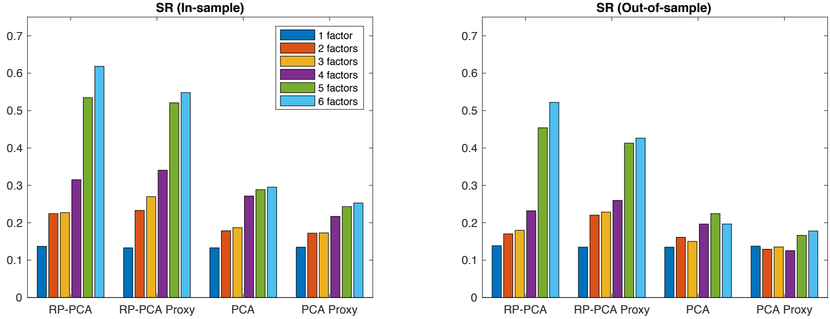


Figure: Maximal Sharpe-ratios.

⇒ Spike in Sharpe-ratio for 5 factors

Single-sorted portfolios: Pricing error

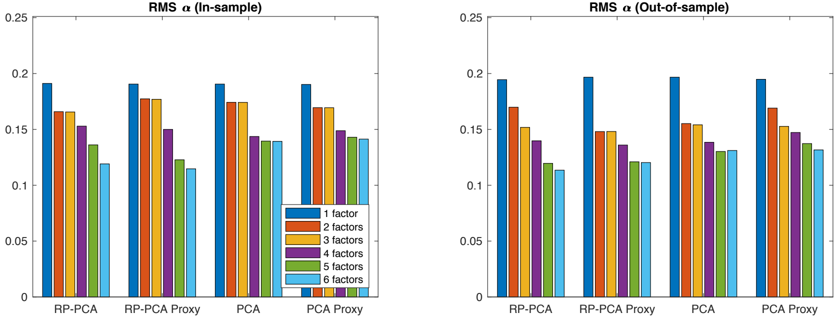


Figure: Root-mean-squared pricing errors.

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

Single-sorted portfolios: Idiosyncratic Variation

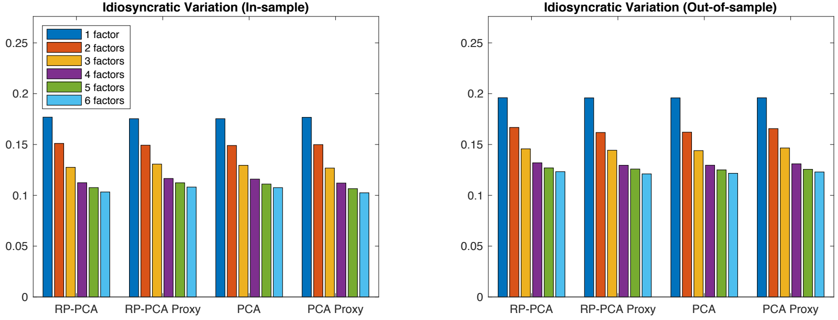


Figure: Unexplained idiosyncratic variation.

⇒ Unexplained variation similar for RP-PCA and PCA

Interpreting factors: 5th proxy factor

5. Proxy RP-PCA	Weights	5. Proxy PCA	Weights
Industry Rel. Reversals (LV) 10	1.12	Leverage 10	1.61
Industry Rel. Reversals (LV) 9	0.98	Value-Profitability 10	1.04
Value-Momentum-Prof. 10	0.95	Asset Turnover 10	1.02
Profitability 10	0.94	Profitability 10	0.99
Industry Mom. Reversals 10	0.91	Asset Turnover 9	0.92
Profitability 2	-0.86	Size 10	0.89
Profitability 3	-0.88	Long Run Reversals 10	0.85
Industry Mom. Reversals 1	-0.90	Sales/Price 10	0.84
Industry Rel. Reversals 2	-0.91	Size 9	0.82
Asset Turnover 1	-0.95	Value-Momentum-Prof. 1	-0.79
Net Operating Assets 1	-0.97	Value-Profitability 1	-0.81
Seasonality 1	-1.00	Profitability 2	-0.81
Value-Profitability 1	-1.12	Profitability 1	-0.89
Short-Term Reversals 1	-1.21	Profitability 4	-0.91
Industry Rel. Reversals (LV) 1	-1.24	Value-Profitability 2	-0.94
Industry Rel. Reversals 1	-1.52	Profitability 3	-1.04
Idiosyncratic Volatility 1	-1.81	Asset Turnover 2	-1.17
Momentum (12m) 1	-1.81	Asset Turnover 1	-1.35

Extreme Deciles

Anomaly	Mean	SD	Sharpe-ratio	Anomaly	Mean	SD	Sharpe-ratio
Accruals - accrual	0.37	3.20	0.12	Momentum (12m) - mom12	1.28	6.91	0.19
Asset Turnover - aturnover	0.40	3.84	0.10	Momentum-Reversals - momrev	0.47	4.82	0.10
Cash Flows/Price - cfp	0.44	4.38	0.10	Net Operating Assets - noa	0.15	5.44	0.03
Composite Issuance - ciss	0.46	3.31	0.14	Price - price	0.03	6.82	0.00
Dividend/Price - divp	0.2	5.11	0.04	Gross Profitability - prof	0.36	3.41	0.11
Earnings/Price - ep	0.57	4.76	0.12	Return on Assets (A) - roaa	0.21	4.07	0.05
Gross Margins - gmargins	0.02	3.34	0.01	Return on Book Equity (A) - roea	0.08	4.40	0.02
Asset Growth - growth	0.33	3.46	0.10	Seasonality - season	0.81	3.94	0.21
Investment Growth - igrowth	0.37	2.69	0.14	Sales Growth - sgrowth	0.05	3.59	0.01
Industry Momentum - indmom	0.49	6.17	0.08	Share Volume - shvol	0.00	6.00	0.00
Industry Mom. Reversals - indmomrev	1.18	3.48	0.34	Size - size	0.29	4.81	0.06
Industry Rel. Reversals - indrrev	1.00	4.11	0.24	Sales/Price sp	0.53	4.26	0.13
Industry Rel. Rev. (L.V.) - indrrevlv	1.34	3.01	0.44	Short-Term Reversals - strev	0.36	5.27	0.07
Investment/Assets - inv	0.49	3.09	0.16	Value-Momentum - valmom	0.51	5.05	0.10
Investment/Capital - invcap	0.13	5.02	0.03	Value-Momentum-Prof. - valmomprof	0.84	4.85	0.17
Idiosyncratic Volatility - ivol	0.56	7.22	0.08	Value-Profitability - valprof	0.76	3.84	0.20
Leverage - lev	0.24	4.58	0.05	Value (A) - value	0.50	4.57	0.11
Long Run Reversals - lrrev	0.46	5.02	0.09	Value (M) - valuem	0.43	5.89	0.07
Momentum (6m) - mom	0.35	6.27	0.06				

Table: Long-Short Portfolios of extreme deciles of 37 single-sorted portfolios from 07/1963 to 12/2017: Mean, standard deviation and Sharpe-ratio.

10 highest SR			10 lowest SR		
Portfolio	Mean	SR	Portfolio	Mean	SR
Ind. Rel. Rev. (L.V.)	1.33	0.44	Return on Assets (A)	0.21	0.05
Industry Mom. Rev.	1.18	0.33	Leverage	0.23	0.05
Industry Rel. Reversals	1.00	0.24	Dividend/Price	0.20	0.03
Seasonality	0.81	0.20	Net Operating Assets	0.15	0.02
Value-Profitability	0.75	0.19	Investment/Capital	0.12	0.02
Momentum (12m)	1.28	0.18	Return on Book Equity (A)	0.08	0.01
Value-Mom-Prof.	0.84	0.17	Gross Margins	0.01	0.00
Investment/Assets	0.48	0.15	Share Volume	0.00	0.00
Composite Issuance	0.45	0.13	Price	0.02	0.00
Investment Growth	0.37	0.13	Sales Growth	0.04	0.00

Extreme Deciles

	In-sample			Out-of-sample		
	SR	RMS α	Idio. Var.	SR	RMS α	Idio. Var.
RP-PCA	**0.57**	**0.17**	**10.40%**	**0.50**	**0.15**	**12.06%**
PCA	0.30	0.22	10.30%	0.24	0.20	11.98%
RP-PCA Proxy	0.58	0.17	10.40%	0.50	0.15	11.97%
PCA Proxy	0.33	0.22	11.09%	0.27	0.18	12.10%
Fama-French 5	0.32	0.30	13.56%	0.31	0.26	13.66%

Table: First and last decile of 37 single-sorted portfolios from 07/1963 to 12/2017 ($N = 74$ and $T = 650$): Maximal Sharpe-ratios, root-mean-squared pricing errors and unexplained idiosyncratic variation. $K = 6$ statistical factors.

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

Interpreting factors: Generalized correlations with proxies

	RP-PCA	PCA
1. Gen. Corr.	1.00	1.00
2. Gen. Corr.	0.99	0.99
3. Gen. Corr.	0.95	0.97
4. Gen. Corr.	0.95	0.94
5. Gen. Corr.	0.71	0.86

Table: Generalized correlations of statistical factors with proxy factors (portfolios of 8 assets).

- Problem in interpreting factors: Factors only identified up to invertible linear transformations.
- Generalized correlations close to 1 measure of how many factors two sets have in common.

⇒ Proxy factors approximate statistical factors.

Extreme Deciles: Maximal Sharpe-ratio

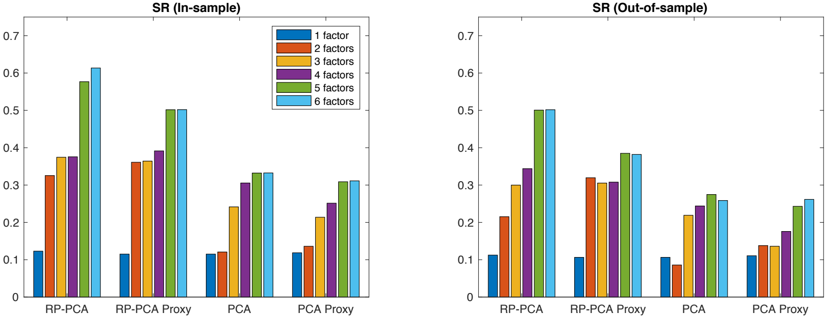


Figure: Maximal Sharpe-ratios.

⇒ Spike in Sharpe-ratio for 5 factors

Extreme Deciles: Pricing error

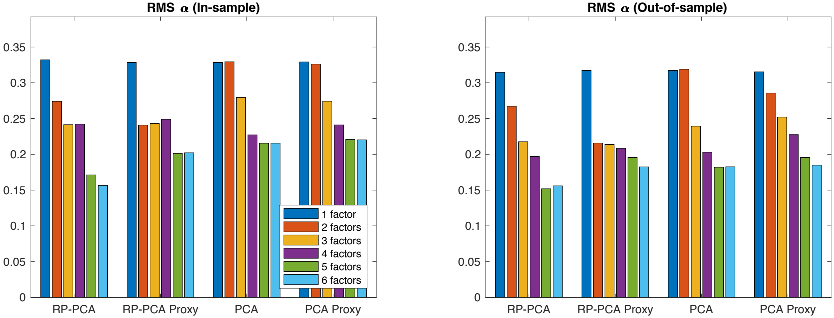


Figure: Root-mean-squared pricing errors.

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

Extreme Deciles: Idiosyncratic Variation

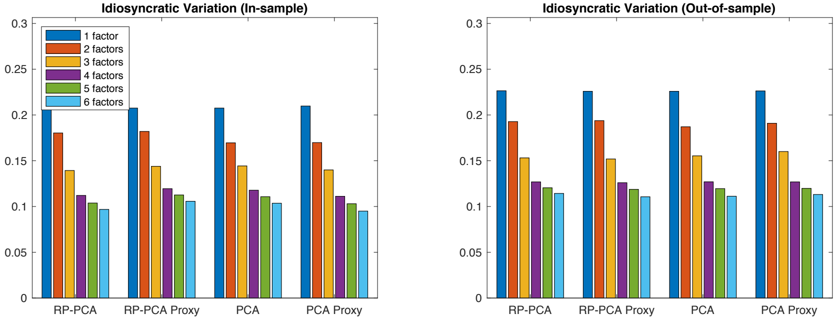


Figure: Unexplained idiosyncratic variation.

⇒ Unexplained variation similar for RP-PCA and PCA

Extreme Deciles: Maximal Sharpe-ratio

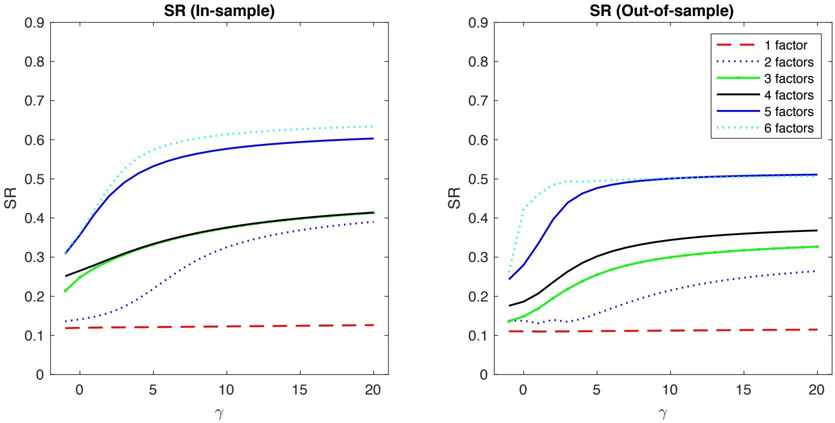


Figure: Maximal Sharpe-ratios for different RP-weights γ and number of factors K

Interpreting factors: 5th proxy factor

5. Proxy RP-PCA	Weights	5. Proxy PCA	Weights
Value 10	1.93	Value-Profitability 10	1.25
Industry Rel. Reversal 10	1.39	Asset Turnover 10	1.15
Price 1	1.31	Profitability 10	0.95
Industry Rel. Reversal (LV) 10	1.26	Sales/Price 10	0.95
Long Run Reversals 10	1.25	Long Run Reversals 10	0.86
Short Run Reversals 1	-1.22	Value-Profitability 1	-0.98
Industry Rel. Reversal (LV) 1	-1.34	Profitability 1	-1.51
Industry Rel. Reversal 1	-1.37	Asset Turnover 1	-1.89

Interpreting factors: Composition of proxies

			RP-PCA				
divp 10	1.53	mom12 10	2.04	size 10	2.14	valuem10	1.93
growth 1	-1.46	mom 10	1.99	ivol 1	2.13	indrrev 10	1.39
igrowth 1	-1.51	indmomrev 10	1.90	valmomprof 10	1.89	price 1	1.31
ep 1	-1.53	mom 1	-2.29	mom12 10	1.84	indrrevlv 10	1.26
invcap 1	-1.69	valuem 10	-2.32	mom 10	1.82	lrrev 10	1.25
shvol 1	-1.72	ivol 1	-2.93	price 1	1.69	strev 1	-1.22
mom12 1	-2.32	price 1	-3.51	shvol 1	1.65	indrrevlv 1	-1.34
ivol 1	-2.48	mom12 1	-4.00	indmomrev 1	-1.57	indrrev 1	-1.37
			PCA				
valuem 10	2.91	divp 10	1.74	indmom 10	2.42	valprof 10	1.25
price 1	2.52	ivol 10	1.69	mom 10	2.39	Aturnover 10	1.15
divp 10	2.26	roea 1	-1.64	valmom 10	2.18	prof 10	0.95
value 10	2.24	mom12 1	-1.65	mom12 10	2.12	sp 10	0.95
lrrev 10	2.06	size 10	-1.82	valmomprof 10	2.12	lrrev 10	0.86
sp 10	1.98	shvol 1	-1.90	indmom 1	-2.38	valprof 1	-0.98
cfp 10	1.92	ivol 1	-3.16	mom12 1	-2.70	prof 1	-1.51
mom12 1	1.88	price 1	-3.21	mom 1	-2.71	Aturnover 1	-1.89

Table: Portfolio-composition of proxy factors for first and last decile of 37 single-sorted portfolios: First proxy factors is an equally-weighted portfolio.

Extreme Deciles: Time-stability of loadings

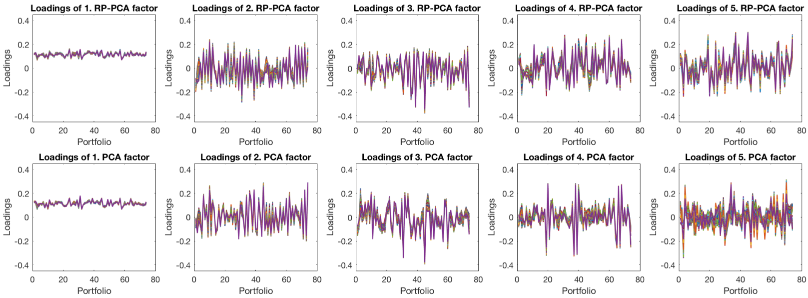


Figure: Time-varying rotated loadings for the first six factors. Loadings are estimated on a rolling window with 240 months.

Extreme Deciles: Time-stability: Generalized correlations

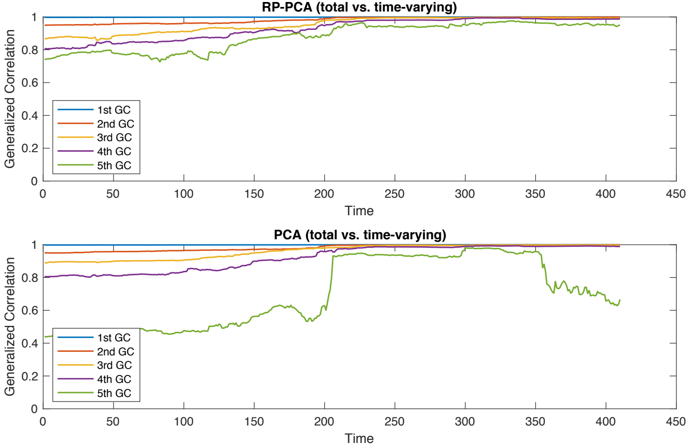


Figure: Generalized correlations between loadings estimated on the whole time horizon $T = 650$ and a rolling window with 240.

Double-sorted portfolios

- Data
  - Monthly return data from 07/1963 to 12/2017 ($T = 650$)
  - 13 double sorted portfolios (consisting of 25 portfolios) from Kenneth French's website
- Factors
  1. **PCA:** $K = 3$
  2. **RP-PCA:** $K = 3$ and $\gamma = 10$
  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	0.23	0.22	0.21	0.13	0.13	0.14
BM and Investment	**0.18**	**0.17**	**0.24**	**0.11**	**0.11**	**0.12**
BM and Profits	0.21	0.20	0.24	0.11	0.12	0.16
Size and Accrual	**0.24**	**0.13**	**0.21**	**0.12**	**0.14**	**0.12**
Size and Beta	0.25	0.24	0.23	0.06	0.07	0.10
Size and Investment	0.29	0.26	0.21	0.11	0.11	0.22
Size and Profits	0.21	0.21	0.25	0.06	0.06	0.16
Size and Momentum	0.21	0.19	0.18	0.15	0.16	0.17
Size and ST-Reversal	0.27	0.25	0.24	0.16	0.17	0.35
Size and Idio. Vol.	0.33	0.31	0.32	0.15	0.16	0.16
Size and Total Vol.	0.32	0.30	0.31	0.16	0.16	0.16
Profits and Invest.	0.26	0.24	0.30	0.11	0.12	0.11
Size and LT-Reversal	0.19	0.18	0.16	0.12	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	0.22	0.18	0.16	0.18	0.19	0.19
BM and Investment	**0.18**	**0.15**	**0.24**	**0.16**	**0.17**	**0.17**
BM and Profits	0.19	0.17	0.23	0.17	0.17	0.19
Size and Accrual	**0.24**	**0.09**	**0.11**	**0.11**	**0.14**	**0.12**
Size and Beta	0.21	0.20	0.16	0.09	0.09	0.09
Size and Investment	0.29	0.23	0.17	0.13	0.14	0.16
Size and Profits	0.21	0.20	0.20	0.10	0.10	0.17
Size and Momentum	0.17	0.12	0.08	0.18	0.18	0.19
Size and ST-Reversal	0.21	0.17	0.23	0.22	0.23	0.25
Size and Idio. Vol.	0.36	0.30	0.28	0.17	0.18	0.18
Size and Total Vol.	0.34	0.28	0.27	0.19	0.20	0.19
Profits and Invest.	0.31	0.25	0.29	0.13	0.15	0.14
Size and LT-Reversal	0.11	0.10	0.04	0.14	0.14	0.14

Maximal Sharpe ratio (Size and accrual)

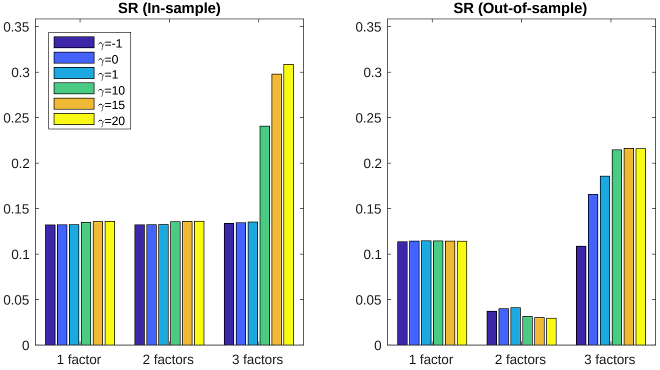


Figure: Maximal Sharpe-ratio by adding factors incrementally.

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

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

Effect of Risk-Premium Weight γ

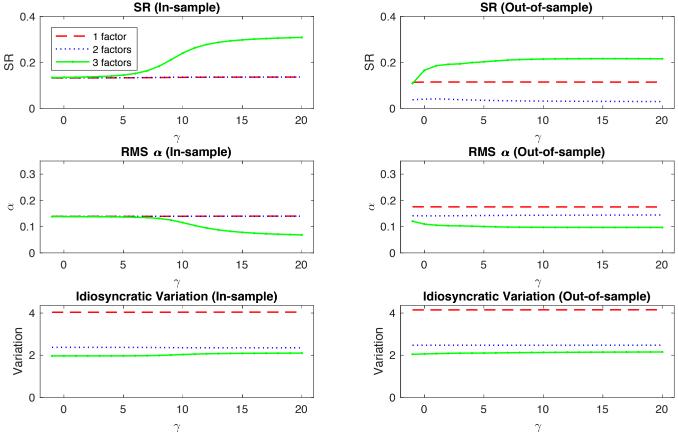


Figure: Maximal Sharpe-ratios, root-mean-squared pricing errors and unexplained idiosyncratic variation.

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

Literature (partial list)

- Large-dimensional factor models with strong factors
  - Bai (2003): Distribution theory
  - Bai and Ng (2017): Robust PCA
  - Fan et al. (2016): Projected PCA for time-varying loadings
  - Kelly et al. (2017): Instrumented PCA for time-varying loadings
  - Pelger (2016), Aït-Sahalia and Xiu (2015): High-frequency
- Large-dimensional factor models with weak factors (based on random matrix theory)
  - Onatski (2012): Phase transition phenomena
  - Benauch-Georges and Nadakuditi (2011): Perturbation of large random matrices
- Asset-pricing factors
  - Feng, Giglio and Xiu (2017): Factor selection with double-selection LASSO
  - Kozak, Nagel and Santosh (2017): Bayesian shrinkage

The Model

Approximate Factor Model

- Observe excess returns of N assets over T time periods:

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

- Matrix notation

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

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

The Model

Approximate Factor Model

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

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

  ⇒ Systematic factors should explain a large portion of the variance

  ⇒ Idiosyncratic risk can be weakly correlated

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

$$E[X_i] = E[F]\Lambda_i^\top$$

  ⇒ Systematic factors should explain the cross-section of expected returns

The Model: Estimation of Latent Factors

Conventional approach: PCA (Principal component analysis)

- Apply PCA to the sample covariance matrix:

$$\frac{1}{T}X^\top X - \bar{X}\bar{X}^\top$$

with $\bar{X}$ = sample mean of asset excess returns

- Eigenvectors of largest eigenvalues estimate loadings $\hat{\Lambda}$.

Much better approach: Risk-Premium PCA (RP-PCA)

- Apply PCA to a covariance matrix with overweighted mean

$$\frac{1}{T}X^\top X + \gamma\bar{X}\bar{X}^\top \qquad \gamma = \text{risk-premium weight}$$

- Eigenvectors of largest eigenvalues estimate loadings $\hat{\Lambda}$.
- $\hat{F}$ estimator for factors: $\hat{F} = \frac{1}{N}X\hat{\Lambda} = X(\hat{\Lambda}^\top\hat{\Lambda})^{-1}\hat{\Lambda}^\top$.

The Model: Objective Function

Conventional PCA: Objective Function

Minimize the unexplained variance:

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

RP-PCA (Risk-Premium PCA): Objective Function

Minimize jointly the unexplained variance and pricing error

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

with $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{t,i}$ and $\bar{F} = \frac{1}{T}\sum_{t=1}^{T} F_t$ and risk-premium weight γ

The Model: Objective function

Variation objective function:

Minimize the unexplained variation:

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

$$= \min_{\Lambda} \frac{1}{NT} trace \left((XM_\Lambda)^\top (XM_\Lambda) \right) \qquad \text{s.t. } F = X(\Lambda^\top \Lambda)^{-1} \Lambda^\top$$

- Projection matrix $M_\Lambda = I_N - \Lambda(\Lambda^\top \Lambda)^{-1} \Lambda^\top$
- Error (non-systematic risk): $e = X - F\Lambda^\top = XM_\Lambda$
- Λ proportional to eigenvectors of the first K largest eigenvalues of $\frac{1}{NT} X^\top X$ minimizes time-series objective function

$\Rightarrow$ Motivation for PCA

The Model: Objective function

Pricing objective function:

Minimize cross-sectional expected pricing error:

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{E}[X_i]-\hat{E}[F]\Lambda_i^\top\right)^2$$
$$=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}X_i^\top\mathbb{1}-\frac{1}{T}\mathbb{1}^\top F\Lambda_i^\top\right)^2$$
$$=\frac{1}{N}trace\left(\left(\frac{1}{T}\mathbb{1}^\top XM_\Lambda\right)\left(\frac{1}{T}\mathbb{1}^\top XM_\Lambda\right)^\top\right) \qquad \text{s.t. } F=X(\Lambda^\top\Lambda)^{-1}\Lambda^\top$$

- $\mathbb{1}$ is vector $T\times 1$ of 1's and thus $\frac{F^\top\mathbb{1}}{T}$ estimates factor mean
- Why not estimate factors with cross-sectional objective function?
  - Factors not identified
  - Spurious factor detection (Bryzgalova (2016))

The Model: Objective function

Combined objective function: Risk-Premium-PCA

$$\min_{\Lambda,F} \frac{1}{NT} trace\left(\left((XM_\Lambda)^\top (XM_\Lambda)\right)\right) + \gamma \frac{1}{N} trace\left(\left(\frac{1}{T}\mathbb{1}^\top XM_\Lambda\right)\left(\frac{1}{T}\mathbb{1}^\top XM_\Lambda\right)^\top\right)$$

$$= \min_{\Lambda} \frac{1}{NT} trace\left(M_\Lambda X^\top \left(I + \frac{\gamma}{T}\mathbb{1}\mathbb{1}^\top\right) XM_\Lambda\right) \qquad \text{s.t. } F = X(\Lambda^\top \Lambda)^{-1}\Lambda^\top$$

- The objective function is minimized by the eigenvectors of the largest eigenvalues of $\frac{1}{NT}X^\top \left(I_T + \frac{\gamma}{T}\mathbb{1}\mathbb{1}^\top\right) X$.
- $\hat{\Lambda}$ estimator for loadings: proportional to eigenvectors of the first K eigenvalues of $\frac{1}{NT}X^\top \left(I_T + \frac{\gamma}{T}\mathbb{1}\mathbb{1}^\top\right) X$
- $\hat{F}$ estimator for factors: $\frac{1}{N}X\hat{\Lambda} = X(\hat{\Lambda}^\top\hat{\Lambda})^{-1}\hat{\Lambda}^\top$.
- Estimator for the common component $C = F\Lambda$ is $\hat{C} = \hat{F}\hat{\Lambda}^\top$

The Model: Objective function

Weighted Combined objective function:

Straightforward extension to weighted objective function:

$$\min_{\Lambda,F} \frac{1}{NT} trace(Q^\top (X - F\Lambda^\top)^\top (X - F\Lambda^\top)Q)$$
$$+ \gamma \frac{1}{N} trace\left(\mathbb{1}^\top (X - F\Lambda^\top)QQ^\top (X - F\Lambda^\top)^\top \mathbb{1}\right)$$
$$= \min_{\Lambda} trace\left(M_\Lambda Q^\top X^\top \left(I + \frac{\gamma}{T}\mathbb{1}\mathbb{1}^\top\right) XQM_\Lambda\right) \qquad \text{s.t. } F = X(\Lambda^\top\Lambda)^{-1}\Lambda^\top$$

- Cross-sectional weighting matrix Q
- Factors and loadings can be estimated by applying PCA to $Q^\top X^\top \left(I + \frac{\gamma}{T}\mathbb{1}\mathbb{1}^\top\right) XQ$.
- Today: Only Q equal to inverse of a diagonal matrix of standard deviations. For $\gamma = -1$ corresponds to PCA of a correlation matrix.
- Optimal choice of Q: GLS type argument

The Model

Strong vs. weak factor models

- Strong factor model ($\frac{1}{N}\Lambda^\top\Lambda$ bounded)
  - Interpretation: 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
  - $\Rightarrow$ optimal γ relatively small
- Weak factor model ($\Lambda^\top\Lambda$ bounded)
  - Interpretation: weak factors affect a smaller fraction of assets
  - Weak factors lead to large but bounded eigenvalues
  - $\Rightarrow$ RP-PCA detects weak factors which cannot be detected by PCA
  - $\Rightarrow$ optimal γ relatively large

Weak Factor Model

Weak Factor Model

- Weak factors either have a small variance or affect a smaller fraction of assets:
- $\Lambda^\top \Lambda$ bounded (after normalizing factor variances)
- Statistical model: Spiked covariance models from random matrix theory
- Eigenvalues of sample covariance matrix separate into two areas:
  - The bulk, majority of eigenvalues
  - The extremes, a few large outliers
- Bulk spectrum converges to generalized Marchenko-Pastur distribution (under certain conditions)

Weak Factor Model

Weak Factor Model

- Large eigenvalues converge either to
  - A biased value characterized by the Stieltjes transform of the bulk spectrum
  - To the bulk of the spectrum if the true eigenvalue is below some critical threshold

  ⇒ Phase transition phenomena: estimated eigenvectors orthogonal to true eigenvectors if eigenvalues too small
- Onatski (2012): Weak factor model with phase transition phenomena
- Problem: All models in the literature assume that random processes have **mean zero**

⇒ RP-PCA implicitly uses non-zero means of random variables

⇒ New tools necessary!

Weak Factor Model

Assumption 1: Weak Factor Model

1. **Rate:** Assume that $\frac{N}{T} \to c$ with $0 < c < \infty$.
2. **Factors:** F are uncorrelated among each other and are independent of e and Λ and have bounded first two moments.

$$\hat{\mu}_F := \frac{1}{T}\sum_{t=1}^{T} F_t \xrightarrow{p} \mu_F \qquad \hat{\Sigma}_F := \frac{1}{T} F_t F_t^\top \xrightarrow{p} \Sigma_F = \begin{pmatrix} \sigma^2_{F_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2_{F_K} \end{pmatrix}$$

3. **Loadings:** The column vectors of the loadings Λ are orthogonally invariant and independent of ϵ and F (e.g. $\Lambda_{i,k} \sim N(0, \frac{1}{N})$) and

$$\Lambda^\top \Lambda = I_K$$

4. **Residuals:** $e = \epsilon\Sigma$ with $\epsilon_{t,i} \sim N(0,1)$. The empirical eigenvalue distribution function of Σ converges to a non-random spectral distribution function with compact support and supremum of support b. Largest eigenvalues of Σ converge to b.

Weak Factor Model

Definition: Weak Factor Model

- Average idiosyncratic noise $\sigma_e^2 := trace(\Sigma)/N$
- Denote by $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$ the ordered eigenvalues of $\frac{1}{T} e^\top e$. The Cauchy transform (also called Stieltjes transform) of the eigenvalues is the almost sure limit:

$$G(z) := a.s. \lim_{T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{z-\lambda_i} = a.s. \lim_{T\to\infty} \frac{1}{N} trace\left(\left(zI_N - \frac{1}{T} e^\top e \right) \right)^{-1}$$

- B-function

$$B(z) := a.s. \lim_{T\to\infty} \frac{c}{N} \sum_{i=1}^{N} \frac{\lambda_i}{(z-\lambda_i)^2}$$
$$= a.s. \lim_{T\to\infty} \frac{c}{N} trace\left(\left(\left(zI_N - \frac{1}{T} e^\top e \right) \right)^{-2} \left(\frac{1}{T} e^\top e \right) \right)$$

Weak Factor Model

Intuition: Weak Factor Model

- "Signal" matrix for PCA of covariance matrix ($\gamma = -1$):

$$\Sigma_F + c\sigma_e^2 I_K$$

K largest eigenvalues $\theta_1^{PCA}, ..., \theta_K^{PCA}$ measure strength of signal

- "Signal" matrix for RP-PCA:

$$\begin{pmatrix} \Sigma_F + c\sigma_e^2 & \Sigma_F^{1/2}\mu_F(1+\tilde{\gamma}) \\ \mu_F^\top \Sigma_F^{1/2}(1+\tilde{\gamma}) & (1+\gamma)(\mu_F^\top \mu_F + c\sigma_e^2) \end{pmatrix} \qquad (1+\tilde{\gamma})^2 = 1+\gamma$$

K largest eigenvalues $\theta_1^{\text{RP-PCA}}, ..., \theta_K^{\text{RP-PCA}}$ measure strength of signal

- RP-PCA signal matrix is "close" to

$$\Sigma_F + (1+\gamma)\mu_F\mu_F^\top + c\sigma_e^2 I_K$$

Weak Factor Model

Theorem 1: Risk-Premium PCA under weak factor model

Assumption 1 holds. The first K largest eigenvalues $\hat{\theta}_i$ $i = 1, ..., K$ of $\frac{1}{T} X^\top \left(I_T + \gamma \frac{\mathbb{1}\mathbb{1}^\top}{T} \right) X$ satisfy

$$\hat{\theta}_i \xrightarrow{p} \begin{cases} G^{-1}\left(\frac{1}{\theta_i}\right) & \text{if } \theta_i > \theta_{crit} = \lim_{z \downarrow b} \frac{1}{G(z)} \\ b & \text{otherwise} \end{cases}$$

The correlation of the estimated with the true factors converges to

$$\widehat{Corr}(F, \hat{F}) \xrightarrow{p} \underbrace{\tilde{U}}_{rotation} \begin{pmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \rho_K \end{pmatrix} \underbrace{\tilde{V}}_{rotation}$$

with

$$\rho_i^2 \xrightarrow{p} \begin{cases} \frac{1}{1+\theta_i B(\hat{\theta}_i))} & \text{if } \theta_i > \theta_{crit} \\ 0 & \text{otherwise} \end{cases}$$

Weak Factor Model

Optimal choice or risk premium weight γ

- Critical value θ_{crit} and function $B(.)$ depend only on the noise distribution and are known in closed-form
- If $\mu_F \neq 0$ and $\gamma > -1$ then RP-PCA signals are always larger than PCA signals:

$$\theta_i^{\text{RP-PCA}} > \theta_i^{PCA}$$

$\Rightarrow$ RP-PCA can detect factors that cannot be detected with PCA

- For $\theta_i > \theta_{crit}$ correlation ρ_i^2 is strictly increasing in θ_i.
- The rotation matrices satisfy $\tilde{U}^\top \tilde{U} \leq I_K$ and $\tilde{V}^\top \tilde{V} \leq I_K$.

$\Rightarrow$ $\widehat{Corr}(F, \hat{F})$ is not necessarily an increasing function in θ.

$\Rightarrow$ Based on closed-form expression choose optimal RP-weight γ

Strong Factor Model

Strong Factor Model

- Strong factors affect most assets: e.g. market factor
- $\frac{1}{N}\Lambda^\top\Lambda$ bounded (after normalizing factor variances)
- Statistical model: Bai and Ng (2002) and Bai (2003) framework
- Factors and loadings can be estimated consistently and are asymptotically normal distributed
- RP-PCA provides a more efficient estimator of the loadings
- Assumptions essentially identical to Bai (2003)

Strong Factor Model

Asymptotic Distribution (up to rotation)

- 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 $W^2 = \left(I_T + \gamma \frac{\mathbb{1}\mathbb{1}^\top}{T}\right)$ 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 (i.e. $\gamma > -1$) always more efficient than PCA
- Optimal γ typically smaller than optimal value from weak factor model
- RP-PCA and PCA are both consistent

Simplified Strong Factor Model

Assumption 2: Simplified Strong Factor Model

1. **Rate:** Same as in Assumption 1
2. **Factors:** Same as in Assumption 1
3. **Loadings:** $\Lambda^\top \Lambda / N \xrightarrow{p} I_K$ and all loadings are bounded.
4. **Residuals:** $e = \epsilon \Sigma$ with $\epsilon_{t,i} \sim N(0,1)$. All elements and all row sums of Σ are bounded.

Simplified Strong Factor Model

Proposition: Simplified Strong Factor Model

Assumption 2 holds. Then:

1. The factors and loadings can be estimated consistently.
2. The asymptotic distribution of the factors is not affected by γ.
3. The asymptotic distribution of the loadings is given by

$$\sqrt{T}\left(H\hat{\Lambda}_i - \Lambda_i\right) \xrightarrow{D} N(0, \Omega_i)$$

$$\begin{aligned}\Omega_i =& \sigma_{e_i}^2 \left(\Sigma_F + (1+\gamma)\mu_F\mu_F^\top\right)^{-1} \left(\Sigma_F + (1+\gamma)^2\mu_F\mu_F^\top\right) \\ & \left(\Sigma_F + (1+\gamma)\mu_F\mu_F^\top\right)^{-1} \\ E[e_{t,i}^2] =& \sigma_{e_i}^2, \quad H \text{ full rank matrix}\end{aligned}$$

4. $\gamma = 0$ is **optimal choice** for smallest asymptotic variance. $\gamma = -1$, i.e. the covariance matrix, is not efficient.

Strong Factor Model

Theorem 2: Strong Factor Model

Assumption 2 holds and $\gamma \in [-1, \infty)$. Then:

- For any choice of γ the factors, loadings and common components can be estimated consistently pointwise.
- If $\frac{\sqrt{T}}{N} \to 0$ then $\sqrt{T}\left(H^\top \hat{\Lambda}_i - \Lambda_i\right) \xrightarrow{D} N(0, \Phi)$

  $$\Phi = \left(\Sigma_F + (\gamma+1)\mu_F\mu_F^\top\right)^{-1}\left(\Omega_{1,1} + \gamma\mu_F\Omega_{2,1} + \gamma\Omega_{1,2}\mu_F + \gamma^2\mu_F\Omega_{2,2}\mu_F\right) \cdot \left(\Sigma_F + (\gamma+1)\mu_F\mu_F^\top\right)^{-1}$$

  For $\gamma = -1$ this simplifies to the conventional case $\Sigma_F^{-1}\Omega_{1,1}\Sigma_F^{-1}$.
- If $\frac{\sqrt{N}}{T} \to 0$ then the asymptotic distribution of the factors is not affected by the choice of γ.
- The asymptotic distribution of the common component depends on γ if and only if $\frac{N}{T}$ does not go to zero. For $\frac{T}{N} \to 0$ $\sqrt{T}\left(\hat{C}_{t,i} - C_{t,i}\right) \xrightarrow{D} N\left(0, F_t^\top \Phi F_t\right)$

Time-varying loadings

Model with time-varying loadings

- Observe panel of excess returns and L covariates $Z_{i,t-1,l}$:

$$X_{t,i} = \underset{1\times K}{F_t}^{\top} \underset{K\times 1}{g}\left(Z_{i,t-1,1}, \ldots, Z_{i,t-1,L}\right) + e_{t,i}$$

- Loadings are function of L covariates $Z_{i,t-1,l}$ with $l = 1, \ldots, L$
  e.g. characteristics like size, book-to-market ratio, past returns, ...
- Factors and loading function are latent
- Idea: Similar to Projected PCA (Fan, Liao and Wang (2016)) and Instrumented PCA (Kelly, Pruitt, Su (2017)), but
  - we include the pricing error penalty
  - allow for general interactions between covariates

Time-varying loadings

Projected RP-PCA (work in progress)

- Approximate nonlinear function $g_k(.)$ by basis functions $\phi_m(.)$:

$$g_k(Z_{i,t-1}) = \sum_{m=1}^{M} b_{m,k} \phi_m(Z_{i,t-1}) \qquad \underbrace{g(Z_{t-1})}_{K \times N} = \underbrace{B^\top}_{K \times M} \underbrace{\Phi(Z_{t-1})}_{M \times N}$$

- Apply RP-PCA to projected data $\tilde{X}_t = X_t \Phi(Z_{t-1})^\top$

$$\tilde{X}_t = F_t B^\top \Phi(Z_{t-1}) \Phi(Z_{t-1})^\top + e_t \Phi(Z_{t-1})^\top = F_t \tilde{B} + \tilde{e}_t$$

- Special case: $\phi_m = \mathbb{1}_{\{Z_{t-1} \in I_m\}} \Rightarrow \tilde{X}$ characteristics sorted portfolios
- Obtain arbitrary interactions and break curse of dimensionality by conditional tree sorting projection
- Intuition: Projection creates M portfolios sorted on any functional form and interaction of covariates Z_{t-1}.

Simulation

Simulation parameters

- Parameters as in the empirical application
- $N = 370$ and $T = 650$.
- Factors:
  - $K = 4$ or $K = 1$
  - Factors $F_t \sim N(\mu_F, \Sigma_F)$
  - $\Sigma_F = diag(5, 0.3, 0.1, \sigma_F^2)$ with $\sigma_F^2 \in \{0.03, 0.05, 0.1\}$
  - $SR_F = (0.12, 0.1, 0.3, sr)$ with $sr \in \{0.8, 0.5, 0.3, 0.2\}$
- Loadings: $\Lambda_i \sim N(0, I_K)$
- Residuals: $e_t \sim \epsilon_t \Sigma$ with empirical correlation matrix and $\sigma_e^2 = 1$.

Simulation

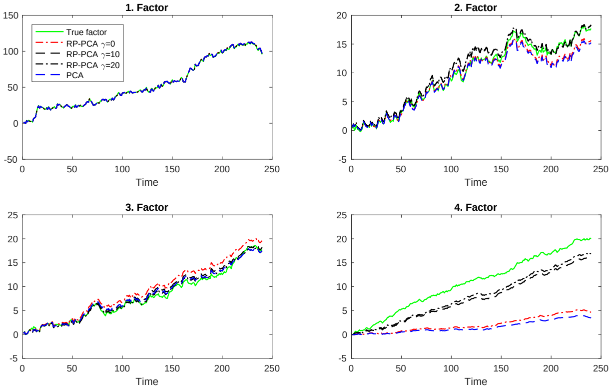


Figure: Sample paths of the cumulative returns of the first four factors and the estimated factor processes.The fourth factor has a variance $\sigma_F^2 = 0.03$ and Sharpe-ratio $sr = 0.5$.

Simulation: Multifactor Model

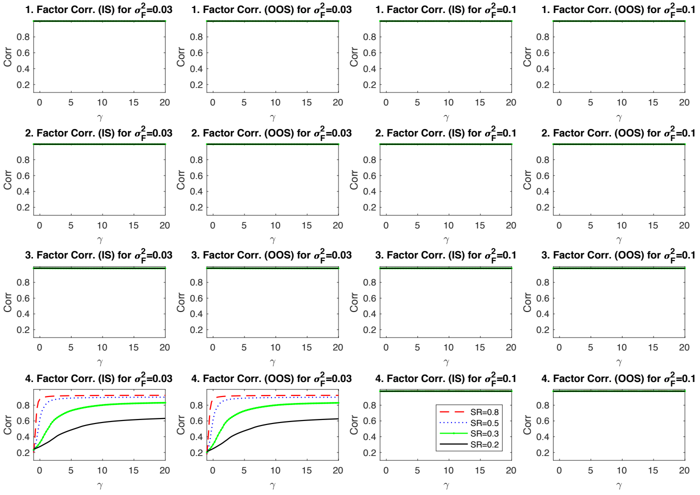


Figure: Correlation of estimated with true factor.

Simulation: Multifactor Model

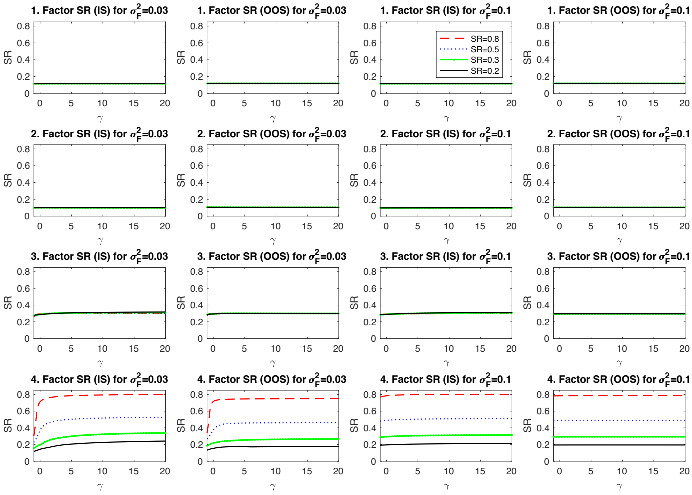
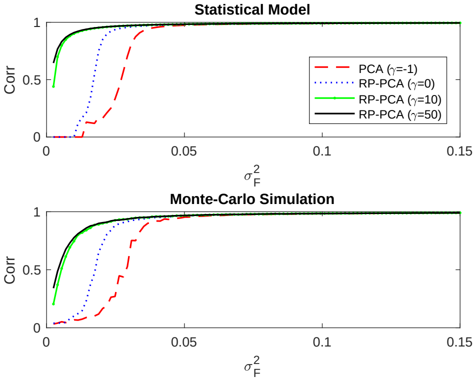


Figure: Maximal Sharpe-ratio of factors.

Simulation: Weak factor model prediction



Correlations between estimated and true factor based on the weak factor model prediction and Monte-Carlo simulations. The Sharpe-ratio of the factor is 0.8. The normalized variance of the factors corresponds to $\sigma_F^2 \cdot N$.

Weak Factor Model: Dependent residuals

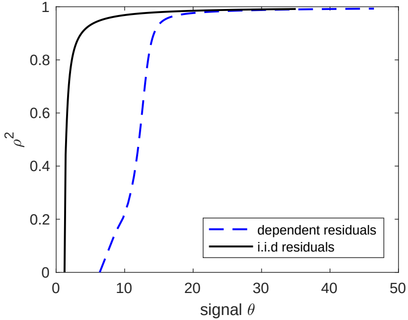
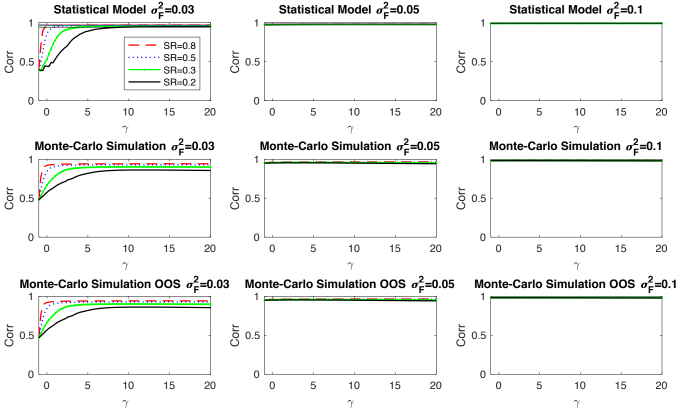


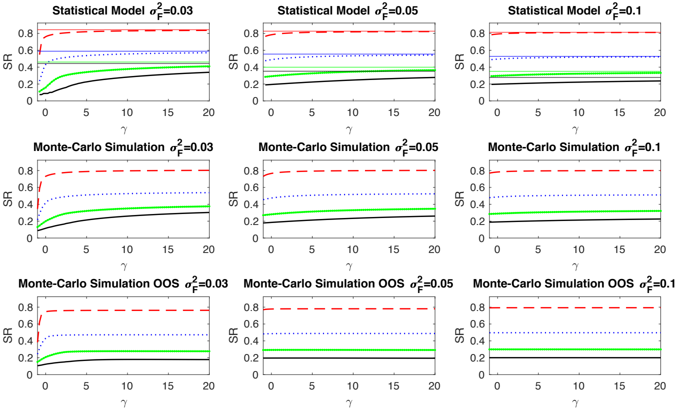
Figure: Model-implied values of ρ_i^2 ($\frac{1}{1+\theta_i B(\hat{\theta}_i))}$ if $\theta_i > \sigma_{crit}^2$ and 0 otherwise) for different signals θ_i. The average noise level is normalized in both cases to $\sigma_e^2 = 1$.

Simulation: Weak factor model prediction



Correlation of estimated with true factors for different variances and Sharpe-ratios of the factor and for different RP-weights γ.

Simulation: Weak factor model prediction



Sharpe-ratio for different variances and Sharpe-ratios of the factor and for different RP-weights γ. The residuals have the empirical residual correlation matrix.