Estimating Asset Pricing Factors from Large-Dimensional Data

Markus Pelger [1] Martin Lettau [2]

[1]Stanford University

[2]UC Berkeley

April 21st, 2017

Machine Learning in Finance Workshop 2017
DSI Columbia University and Bloomberg LP

Motivation: Asset Pricing with Risk Factors

The Challenge of Asset Pricing

- Most important question in finance: Why are prices different for different assets?
- Fundamental insight: Arbitrage Pricing Theory: Prices of financial assets should be explained by systematic risk factors.
- Problem: “Chaos” in asset pricing factors: Over 330 potential asset pricing factors published!
- Fundamental question: Which factors are really important in explaining expected returns? Which are subsumed by others?

Goals of this paper:

Bring order into “factor chaos”

$\Rightarrow$ Summarize the pricing information of a large number of assets in a small number of factors

Why is it important?

Importance of finding the "right" factors

- Risk-management
  - Factors explain risk-return trade-off
  - Factors allows to manage systematic risk exposure
- Investment decisions
  - Find investment decision with high risk-adjusted expected returns
  - "Smart beta" investments = investment strategies based on factors
- Arbitrage opportunities
  - Find underpriced assets and earn "alpha"

Contribution of this paper

Contribution

- 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. Allows time-variation in factor structure

Contribution of this paper

Results

- 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)

The Model

Approximate Factor Model

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

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

- 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):

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

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

  $\Rightarrow$ Idiosyncratic risk can be weakly correlated

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

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

  $\Rightarrow$ 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

Interpretation of Risk-Premium-PCA (**RP-PCA**):

1. Combines 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. Information interpretation: (GMM interpretation)
   - PCA of a covariance matrix uses only the second moment but ignores first moment
   - Using more information leads to more efficient estimates. RP-PCA combines first and second moments efficiently.

The Model

Interpretation of Risk-Premium-PCA (**RP-PCA**): continued

4. Signal-strengthening: Intuitively the matrix $\frac{1}{T}X^\top X + \gamma \bar{X}\bar{X}^\top$ converges to

$$\Lambda \left(\Sigma_F + (1+\gamma)\mu_F \mu_F^\top \right) \Lambda^\top + Var(e)$$

with $\Sigma_F = Var(F)$ and $\mu_F = E[F]$. The signal of weak factors with a small variance can be "pushed up" by their mean with the right γ.

Illustration

Illustration: Anomaly-sorted portfolios (Size and accrual)

- Factors
  1. **PCA:** Estimation based on PCA of correlation matrix, $K = 3$
  2. **RP-PCA:** $K = 3$ and $\gamma = 100$
  3. **Fama-French 5** factor model: market, size, value, profitability and investment
  4. **Specific** factors: market, size and accrual
- Data
  - Double-sorted portfolios according to size and accrual (from Kenneth French's website)
  - Monthly return data from July 1963 to December 2013 (T = 606) for N = 25 portfolios

Portfolio Data: In-sample (Size and accrual)

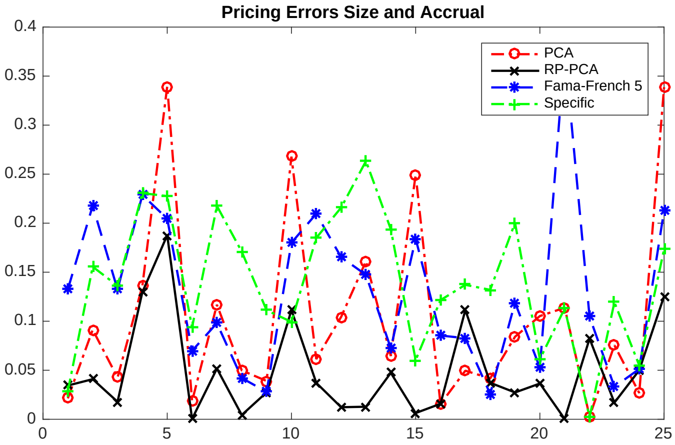
	SR	RMS α	Fama-MacBeth
RP-PCA	**0.305**	**0.068**	**44.570**
PCA	0.135	0.141	89.946
Fama-French	0.344	0.154	61.979
Specific	0.173	0.155	76.041

Table: Maximal Sharpe-ratios, root-mean-squared pricing errors and Fama-MacBeth test statistics. $K = 3$ statistical factors and risk-premium weight $\gamma = 100$.

- SR: Maximum Sharpe-ratio of linear combination of factors
- Cross-sectional pricing errors α:
  - Pricing error $\alpha_i = E[X_i] - E[F]\Lambda_i^\top$
  - RMS α: Root-mean-squared pricing errors $\sqrt{\frac{1}{N}\sum_{i=1}^{N} {\alpha_i}^2}$

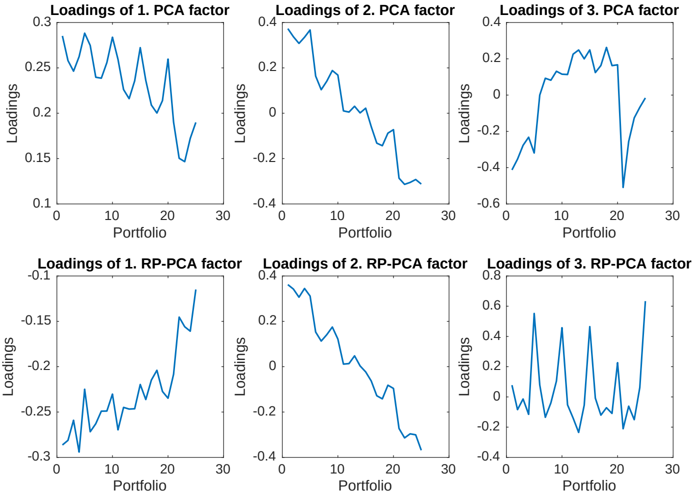
$\Rightarrow$ RP-PCA significantly better than PCA and quantile-sorted factors.

Cross-sectional α's for sorted portfolios (Size and Accrual)



$\Rightarrow$ RP-PCA avoids large pricing errors due to penalty term.

Loadings for statistical factors (Size and Accrual)



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

Maximal Incremental Sharpe Ratio

	PCA	RP-PCA
1 Factor	0.134	0.137
2 Factors	0.135	0.139
3 Factors	0.135	0.305

Table: Maximal Sharpe-ratio by adding factors incrementally. For the statistical factor estimators we use $K = 3$ factors and $\gamma = 100$.

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

⇒ Better performance of RP-PCA because of third accrual factor.

Portfolio Data: Objective function (Size and Accrual)

	PCA TS	RP-PCA TS	PCA XS	RP-PCA XS
1 Factor	3.308	3.617	0.014	0.002
2 Factors	1.937	2.240	0.014	0.002
3 Factors	1.623	1.751	0.014	0.000

Table: Time-series and cross-sectional objective functions.

$\Rightarrow$ RP-PCA and PCA explain the same amount of variation.

$\Rightarrow$ PR-PCA explains cross-sectional pricing much better.

$\Rightarrow$ Motivation for risk-premium weight $\gamma = 100$.

Portfolio Data: Out-of-sample (Size and Accrual)

	Out-of-sample	In-sample
RP-PCA	**0.097**	**0.090**
PCA	0.128	0.146
Fama-French 5	0.111	0.102
Specific	0.134	0.126

Table: Root-mean-squared pricing errors. Out-of-sample factors are estimated with a rolling window. $K = 3$ statistical factors and risk-premium weight $\gamma = 100$.

$\Rightarrow$ RP-PCA performs better in- and out-of-sample.

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

  $\Rightarrow$ 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, e.g. value factor

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

  $\Rightarrow$ optimal γ relatively large

The Model

Strong vs. weak factor models

- Consequences for eigenvalues of $\frac{1}{T}X^\top X$:
  - Strong factors lead to exploding eigenvalues
  - Weak factors lead to large but bounded eigenvalues
- Empirical evidence (equity data): Strong and weak factors:
  - 1st eigenvalue typically substantially larger than rest of spectrum (usually 10 x larger than the 2nd)
  - 2nd and 3rd eigenvalues typically stand out, but similar magnitudes as the rest of the spectrum

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

Intuition: Weak Factor Model

- "Signal" matrix for PCA of covariance matrix:

$$\Lambda \Sigma_F \Lambda^\top$$

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

- "Signal" matrix for RP-PCA:

$$\Lambda \left(\Sigma_F + (1+\gamma)\mu_F \mu_F^\top \right) \Lambda^\top$$

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

- If $\mu_F \neq 0$ and $\gamma > 1$ then RP-PCA signal always larger than PCA signal:

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

Weak Factor Model

Theorem 1: Risk-Premium PCA under weak factor model

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

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

- Critical value θ_{crit} and function $B(.)$ depend only on the noise distribution and are known in closed-form
- Based on closed-form expression choose optimal RP-weight γ
- For $\theta_i > \theta_{crit}$ ρ_i^2 is strictly increasing in θ_i.

$\Rightarrow$ RP-PCA strictly dominates PCA

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):
  - Asymptotically $\hat{\Lambda}$ behaves like OLS regression of F on X.
  - Asymptotically $\hat{F}$ behaves like OLS regression of Λ on X.
- 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

Time-varying loadings

Model with time-varying loadings

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

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

- Loadings are function of L covariates $Z_{i,t-1,l}$ with $l = 1, \dots, L$
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 (2009)), 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 sieve method

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

- Appropriate basis functions $\phi_1(.), ..., \phi_M(.)$ (e.g. splines, kernels)
- Apply RP-PCA to projected data $\tilde{X}_t = X_t\Phi(Z_{t-1})^\top$

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

- 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}.

Portfolio Data

Portfolio Data

- Data
  - Monthly return data from July 1963 to December 2013 ($T = 606$)
  - 13 double sorted portfolios (consisting of 25 portfolios) from Kenneth French's website and 49 industry portfolios
- Factors
  1. **PCA:** $K = 3$
  2. **RP-PCA:** $K = 3$ and $\gamma = 100$
  3. **Fama-French 5** factor model: market, size, value, profitability and investment
  4. **Specific** factors: market + two specific anomaly long-short factors

Pricing errors α (in-sample)

	RP-PCA	PCA	FF 5	Specific
Size and BM	0.13	0.14	0.12	0.20
BM and Investment	**0.07**	**0.12**	**0.14**	**0.13**
BM and Operating Profits	0.11	0.12	0.14	0.17
Size and Accrual	**0.07**	**0.14**	**0.15**	**0.16**
Size and Beta	0.06	0.07	0.08	0.17
Size and Investment	0.11	0.13	0.11	0.20
Size and Operating Profits	0.06	0.07	0.08	0.16
Size and Short-Term Reversal	0.15	0.16	0.24	0.33
Size and Long-Term Reversal	0.11	0.13	0.09	0.20
Size and Res. Var.	0.17	0.18	0.21	0.22
Size and Total Var.	0.18	0.19	0.22	0.21
Operating Profits and Investment	**0.11**	**0.14**	**0.12**	**0.14**
Size and Net Share Iss.	**0.14**	**0.16**	**0.13**	**0.17**
49 Industries	0.14	0.16	0.13	0.29

Pricing errors α (out-of-sample)

	RP-PCA	PCA	FF 5	Specific
Size and BM	0.17	0.19	0.14	0.21
BM and Investment	**0.12**	**0.16**	**0.11**	**0.14**
BM and Operating Profits	0.15	0.18	0.15	0.17
Size and Accrual	**0.10**	**0.13**	**0.11**	**0.13**
Size and Beta	0.09	0.10	0.07	0.09
Size and Investment	0.14	0.17	0.12	0.19
Size and Operating Profits	0.09	0.12	0.09	0.18
Size and Short-Term Reversal	0.17	0.19	0.09	0.18
Size and Long-Term Reversal	0.13	0.14	0.09	0.14
Size and Res. Var.	0.17	0.20	0.18	0.26
Size and Total Var.	0.17	0.21	0.20	0.26
Operating Profits and Investment	**0.13**	**0.17**	**0.13**	**0.16**
Size and Net Share Iss.	**0.14**	**0.21**	**0.16**	**0.18**
49 Industries	0.26	0.24	0.21	0.25

Maximum Sharpe-Ratios

	RP-PCA	PCA	Specific
Size and BM	0.25	0.22	0.16
BM and Investment	**0.26**	**0.17**	**0.24**
BM and Operating Profits	0.24	0.22	0.25
Size and Accrual	**0.30**	**0.13**	**0.17**
Size and Beta	0.23	0.23	0.17
Size and Investment	0.30	0.26	0.23
Size and Operating Profits	0.22	0.21	0.18
Size and Short-Term Reversal	0.26	0.20	0.25
Size and Long-Term Reversal	0.23	0.18	0.15
Size and Res. Var.	0.33	0.30	0.32
Size and Total Var.	0.32	0.28	0.32
Operating Profits and Investment	**0.31**	**0.24**	**0.34**
Size and Net Share Iss.	**0.33**	**0.25**	**0.35**
49 Industries	0.35	0.25	0.11

Portfolio Data

Portfolio Data

- Monthly return data from July 1963 to December 2013 ($T = 606$) for $N = 199$ portfolios
  - Novy-Marx and Velikov (2014) data: 150 portfolios sorted according to 15 anomalies (same data as in Kozak, Nagel and Santosh (2015))
  - 49 industry portfolios from Kenneth French's website

  1. **Fama-French 5:** The five factor model of Fama-French (market, size, value, investment and operating profitability, all from Kenneth French's website).
  2. **Specific**: Market, value, value-momementum-profitibility and volatility factors.

- Number of statistical factors $K = 4$ and $\gamma = 100$.

Portfolio Data I: 15 Novy-Marx factors and portfolios

- Size
- Gross Profitability
- Value
- Value Prof
- Accruals
- Net Issuance
- Asset Growth
- Investment
- Piotrotski F-Score
- ValMomProf
- ValMom
- Idiosyncratic Vol
- Momentum
- Long Run Reversal
- Beta Arbitrage.

Portfolio Data: In-sample

	SR	RMS α	Fama-MacBeth
RP-PCA	**0.417**	**0.135**	729.944
PCA	0.155	0.213	820.804
Fama-French	0.344	0.225	801.013
Specific	0.413	0.152	731.392

Table: Maximal Sharpe-ratios, root-mean-squared pricing errors and Fama-MacBeth test statistics. $K = 4$ statistical factors and risk-premium weight $\gamma = 100$.

- RP-PCA strongly dominates PCA and Fama-French 5 factors
- Specific factors (Market, Value, Value-Momementum-Profitibility and Volatility) perform similar to RP-PCA.

Portfolio Data: Out-of-sample

	Out-of-sample	In-sample
RP-PCA	0.178	0.145
PCA	0.202	0.208
Fama-French 5	0.182	0.182
Specific	0.154	0.137

Table: Root-mean-squared pricing errors. Out-of-sample factors are estimated with a rolling window. $K = 4$ statistical factors and risk-premium weight $\gamma = 100$.

$\Rightarrow$ RP-PCA performs well in- and out-of-sample.

Portfolio Data: Interpreting factors

	PCA	RP-PCA
1. Gen. Corr.	0.997	0.997
2. Gen. Corr.	0.898	0.925
3. Gen. Corr.	0.809	0.888
4. Gen. Corr.	0.032	0.741

Table: Generalized Correlations between specific factors and statistical factors.

- 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.
- Specific factors: Market, Value, Value-Momementum-Profitability and Volatility factors.

⇒ Specific factors approximate RP-PCA factors.

Conclusion

Methodology

- New estimator for estimating priced latent factors from large data sets
- Combines time-series and cross-sectional 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 estimation based on PCA of the covariance matrix
- Potential to provide benchmark factors for horse races.
- Promising empirical results.

Literature (partial list)

- Large-dimensional factor models with strong factors
  - Bai (2003): Distribution theory
  - Ahn and Horenstein (2013), Onatski (2010), Bai and Ng (2002): Determining the number of factors
  - Fan et al. (2013): Sparse matrices in factor modeling
  - 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
  - Harvey and Liu (2015): Lucky factors
  - Clarke (2015): Level, slope and curvature for stocks
  - Kozak, Nagel and Santosh (2015): PCA based factors
  - Bryzgalova (2016): Spurious factors

The Model

Time-series objective function:

Minimize the unexplained variance:

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

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

- 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

Cross-sectional objective function:

Minimize cross-sectional expected pricing error:

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

- $\mathbb{1}$ 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

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

Weighted Combined objective function:

Straightforward extension to weighted objective function:

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

- 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

Weak Factor Model

Assumption 1: Weak Factor Model

1. Residual matrix can be represented as $e = \epsilon\Sigma$ 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. The supremum of the support is b.
2. The 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. 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. Assume that $\frac{N}{T} \to c$ with $0 < c < \infty$.

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

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

Weak Factor Model

Estimator

- Risk-premium PCA (RP-PCA): Apply PCA estimation to $S_\gamma := \frac{1}{T} X^\top \left(I_T + \gamma \frac{\mathbb{1}\mathbb{1}^\top}{T} \right) X$
- PCA : Apply PCA to estimated covariance matrix $S_{-1} := \frac{1}{T} X^\top \left(I_T - \frac{\mathbb{1}\mathbb{1}^\top}{T} \right) X$, i.e. $\gamma = -1$.

$\Rightarrow$ PCA special case of RP-PCA

"Signal" Matrix for Covariance PCA

$$M_{Var} = \Sigma_F + c\sigma_e^2 I_K = \begin{pmatrix} \sigma_{F_1}^2 + c\sigma_e^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{F_K}^2 + c\sigma_e^2 \end{pmatrix}$$

$\Rightarrow$ Intuition: Largest K "true" eigenvalues of S_{-1}.

Weak Factor Model

Lemma: Covariance PCA

Assumption 1 holds. Define the critical value $\sigma^2_{crit} = \lim_{z \downarrow b} \frac{1}{G(z)}$. The first K largest eigenvalues $\hat{\lambda}_i$ of S_{-1} satisfy for $i = 1, ..., K$

$$\hat{\lambda}_i \xrightarrow{p} \begin{cases} G^{-1}\left(\frac{1}{\sigma^2_{F_i} + c\sigma^2_e}\right) & \text{if } \sigma^2_{F_i} + c\sigma^2_e > \sigma^2_{crit} \\ b & \text{otherwise} \end{cases}$$

The correlation between the estimated and true factors converges to

$$\widehat{Corr}(F, \hat{F}) \xrightarrow{p} \begin{pmatrix} \varrho_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \varrho_K \end{pmatrix}$$

with

$$\varrho^2_i \xrightarrow{p} \begin{cases} \frac{1}{1 + (\sigma^2_{F_i} + c\sigma^2_e) B(\hat{\lambda}_i))} & \text{if } \sigma^2_{F_i} + c\sigma^2_e > \sigma^2_{crit} \\ 0 & \text{otherwise} \end{cases}$$

Weak Factor Model

Corollary: Covariance PCA for i.i.d. errors

Assumption 1 holds, $c \geq 1$ and $e_{t,i}$ i.i.d. $N(0, \sigma_e^2)$. The largest K eigenvalues of S_{-1} have the following limiting values:

$$\hat{\lambda}_i \xrightarrow{p} \begin{cases} \sigma_{F_i}^2 + \frac{\sigma_e^2}{\sigma_{F_i}^2}(c + 1 + \sigma_e^2) & \text{if } \sigma_{F_i}^2 + c\sigma_e^2 > \sigma_{crit}^2 \Leftrightarrow \sigma_F^2 > \sqrt{c}\sigma_e^2 \\ \sigma_e^2(1 + \sqrt{c})^2 & \text{otherwise} \end{cases}$$

The correlation between the estimated and true factors converges to

$$\widehat{Corr}(F, \hat{F}) \xrightarrow{p} \begin{pmatrix} \varrho_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \varrho_K \end{pmatrix}$$

with

$$\varrho_i^2 \xrightarrow{p} \begin{cases} \dfrac{1 - \frac{c\sigma_e^4}{\sigma_{F_i}^4}}{1 + \frac{c\sigma_e^2}{\sigma_{F_i}^2} + \frac{\sigma_e^4}{\sigma_{F_i}^4}(c^2 - c)} & \text{if } \sigma_{F_i}^2 + c\sigma_e^2 > \sigma_{crit}^2 \\ 0 & \text{otherwise} \end{cases}$$

Weak Factor Model

"Signal" Matrix for RP-PCA

- "Signal" Matrix for RP-PCA

$$M_{RP} = \begin{pmatrix} \Sigma_F + c\sigma_e^2 & \Sigma_F^{1/2}\mu_F(1+\tilde{\gamma}) \\ \mu_F^\top \Sigma_F^{1/2}(1+\tilde{\gamma}) & (1+\gamma)(\mu_F^\top \mu_F + c\sigma_2^2) \end{pmatrix}$$

Define $\tilde{\gamma} = \sqrt{\gamma+1} - 1$ and note that $(1+\tilde{\gamma})^2 = 1+\gamma$.

$\Rightarrow$ Projection on K demeaned factors and on mean operator.

- Denote by $\theta_1 \geq ... \geq \theta_{K+1}$ the eigenvalues of the "signal matrix" M_{RP} and by $\tilde{U}$ the corresponding orthonormal eigenvectors :

$$\tilde{U}^\top M_{RP} \tilde{U} = \begin{pmatrix} \theta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_{K+1} \end{pmatrix}$$

$\Rightarrow$ Intuition: $\theta_1, ..., \theta_{K+1}$ largest $K+1$ "true" eigenvalues of S_γ.

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 S_γ satisfy

$$\hat{\theta}_i \xrightarrow{p} \begin{cases} G^{-1}\left(\frac{1}{\theta_i}\right) & \text{if } \theta_i > \sigma^2_{crit} = \lim_{z \downarrow b} \frac{1}{G(z)} \\ b & \text{otherwise} \end{cases}$$

The correlation of the estimated with the true factors converges to

$$\widehat{Corr}(F, \hat{F}) \xrightarrow{p} \underbrace{\begin{pmatrix} I_K & 0 \end{pmatrix} \tilde{U}}_{rotation} \begin{pmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \rho_K \\ 0 & \cdots & & 0 \end{pmatrix} \underbrace{D_K^{1/2} \hat{\Sigma}_{\hat{F}}^{-1/2}}_{rotation}$$

with

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

Weak Factor Model

Theorem 1: continued

$$\hat{\Sigma}_{\hat{F}} = D_K^{1/2}\left(\begin{pmatrix}\rho_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \rho_K\\ 0 & \cdots & 0\end{pmatrix}^\top \tilde{U}^\top \begin{pmatrix} I_K & 0\\ 0 & 0\end{pmatrix} \tilde{U} \begin{pmatrix}\rho_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \rho_K\\ 0 & \cdots & 0\end{pmatrix}\right.$$

$$\left. + \begin{pmatrix} 1-\rho_1^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1-\rho_K^2\end{pmatrix}\right) D_K^{1/2}$$

$$D_K = diag\left(\begin{pmatrix}\hat{\theta}_1 & \cdots & \hat{\theta}_K\end{pmatrix}\right)$$

Weak Factor Model

Lemma: Detection of weak factors

If $\gamma > -1$ and $\mu_F \neq 0$, then the first K eigenvalues of M_{RP} are strictly larger than the first K eigenvalues of M_{Var}, i.e.

$$\theta_i > \sigma_{F_i}^2 + c\sigma_e^2$$

For $\theta_i > \sigma_{crit}^2$ it holds that

$$\frac{\partial \hat{\theta}_i}{\partial \theta_i} > 0 \qquad \frac{\partial \rho_i}{\partial \theta_i} > 0 \qquad i = 1, ..., K$$

Thus, if $\gamma > -1$ and $\mu_F \neq 0$, then $\rho_i > \varrho_i$.

$\Rightarrow$ For $\mu_F \neq 0$ RP-PCA always better than PCA.

Weak Factor Model

Example: One-factor model

Assume that there is only one factor, i.e. $K = 1$. The "signal matrix" M_{RP} simplifies to

$$M_{RP} = \begin{pmatrix} \sigma_F^2 + c\sigma_e^2 & \sigma_F \mu(1+\tilde{\gamma}) \\ \mu\sigma_F(1+\tilde{\gamma}) & (\mu^2 + c\sigma_e^2)(1+\gamma) \end{pmatrix}$$

and has the eigenvalues:

$$\begin{aligned} \theta_{1,2} =& \frac{1}{2}\sigma_F^2 + c\sigma_e^2 + (\mu^2 + c\sigma_e^2)(1+\gamma) \\ & \pm \frac{1}{2}\sqrt{(\sigma_F^2 + c\sigma_e^2 + (\mu^2 + c\sigma_e^2)(1+\gamma))^2 - 4(1+\gamma)c\sigma_e^2(\sigma_F^2 + \mu^2 + c\sigma_e^2)} \end{aligned}$$

The eigenvector of first eigenvalue θ_1 has the components

$$\tilde{U}_{1,1} = \frac{\mu\sigma_F(1+\tilde{\gamma})}{\sqrt{(\theta_1 - (\sigma_F^2 + c\sigma_e^2))^2 + \mu^2\sigma_F^2(1+\gamma)}}$$

$$\tilde{U}_{1,2} = \frac{\theta_1 - \sigma_F^2 + c\sigma_e^2}{\sqrt{(\theta_1 - (\sigma_F^2 + c\sigma_e^2))^2 + \mu^2\sigma_F^2(1+\gamma)}}$$

Weak Factor Model

Corollary: One-factor model

The correlation between the estimated and true factor has the following limit:

$$\widehat{Corr}(F, \hat{F}) \xrightarrow{p} \frac{\rho_1}{\sqrt{\rho_1^2 + (1-\rho_1^2)\frac{(\theta_1 - (\sigma_F^2 + c\sigma_e^2))^2 + 1}{\mu^2 \sigma_F^2 (1+\gamma)}}}$$

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):
  - Asymptotically $\hat{\Lambda}$ behaves like OLS regression of F on X.
  - Asymptotically $\hat{F}$ behaves like OLS regression of Λ on X.
- RP-PCA under slightly stronger assumptions as in Bai (2003):
  - Asymptotically $\hat{\Lambda}$ behaves like OLS regression of FW on XW with $W^2 = \left(I_T + \gamma \frac{\mathbb{1}\mathbb{1}^\top}{T}\right)$.
  - Asymptotically $\hat{F}$ behaves like OLS regression of Λ on X.

Asymptotic Expansion

Asymptotic expansions (under slightly stronger assumptions as in Bai (2003)):

1. $\sqrt{T}\left(H^\top \hat{\Lambda}_i - \Lambda_i\right) = \left(\frac{1}{T} F^\top W^2 F\right)^{-1} \frac{1}{\sqrt{T}} F^\top W^2 e_i + O_p\left(\frac{\sqrt{T}}{N}\right) + o_p(1)$
2. $\sqrt{N}\left(H^{\top^{-1}} \hat{F}_t - F_t\right) = \left(\frac{1}{N}\Lambda^\top \Lambda\right)^{-1} \frac{1}{\sqrt{N}} \Lambda^\top e_t^\top + O_p\left(\frac{\sqrt{N}}{T}\right) + o_p(1)$

with known rotation matrix H.

Strong Factor Model

Assumption 2: Strong Factor Model

Assume the same assumptions as in Bai (2003) (Assumption A-G) hold and in addition

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{t,i} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_{t,i} \end{pmatrix} \xrightarrow{D} N(0, \Omega) \qquad \Omega = \begin{pmatrix} \Omega_{1,1} & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{pmatrix}$$

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 $\frac{\sqrt{N}}{T} \to 0$ 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 $\Sigma_F^{-1}\Omega_{1,1}\Sigma_F^{-1}$.
- 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 $\frac{T}{N} \to 0$
  $\sqrt{T}\left(\hat{C}_{t,i} - C_{t,i}\right) \xrightarrow{D} N\left(0, F_t^\top \Phi F_t\right)$

Strong Factor Model

Example 2: Toy model with i.i.d. residuals and $K = 1$

Assume $K = 1$ and $e_{t,i} \overset{i.i.d.}{\sim} (0, \sigma_e^2)$. If Assumption 2 holds and $\frac{\sqrt{T}}{N} \to 0$, then

$$\sqrt{T}\left(\hat{\Lambda}_i - \Lambda_i\right) \overset{D}{\to} N(0, \Omega)$$

with

$$\Omega = \sigma_e^2 \frac{\left(\sigma_F^2 + \mu_F^2(1+\gamma)^2\right)}{\left(\sigma_F^2 + \mu_F^2(1+\gamma)\right)^2}$$

- ⇒ Optimal choice minimizing the asymptotic variance is risk-premium weight $\gamma = 0$.
- ⇒ Choosing $\gamma = -1$, i.e. the covariance matrix for factor estimation, is not efficient.

Simulation

Simulation parameters

- $N = 250$ and $T = 350$.
- Factors: $K = 4$
  - 1. Factor represent the market with $N(1.2, 9)$: Sharpe-ratio of 0.4
  - 2. Factor represents an industry factors following $N(0.1, 1)$: Sharpe-ratio of 0.1.
  - 3. Factor follows $N(0.4, 1)$: Sharpe-ratio of 0.4.
  - 4. Factor has a small variance but high Sharpe-ratio. It follows $N(0.4, 0.16)$: Sharpe-ratio of 1.
- Loadings normalized such that $\frac{1}{N}\Lambda^\top\Lambda$.
  $\Lambda_{i,1} = 1$ and $\Lambda_{i,k} \sim N(0,1)$ for $k = 2, 3, 4$.
- Errors: Cross-sectional and time-series correlation and heteroskedasticity in the residuals. Half of the variation due to non-systematic risk.

Simulation parameters

Errors

Residuals are modeled as $e = \sigma_e D_T A_T \epsilon A_N D_N$:

- ϵ is a $T \times N$ matrix and follows a multivariate standard normal distribution
- Time-series correlation in errors: A_T creates an AR(1) model with parameter $\rho = 0.1$
- Cross-sectional correlation in errors: A_N is a Toeplitz-matrix with $(\beta, \beta, \beta, \beta^2)$ on the right four off-diagonals with $\beta = 0.7$
- Cross-sectional heteroskedasticity: D_N is a diagonal matrix with independent elements following $N(1, 0.2)$
- Time-series heteroskedasticity: D_T is a diagonal matrix with independent elements following $N(1, 0.2)$
- Signal-to-noise ratio: $\sigma_e^2 = 10$
- Parameters produce eigenvalues that are consistent with the data.

Simulation

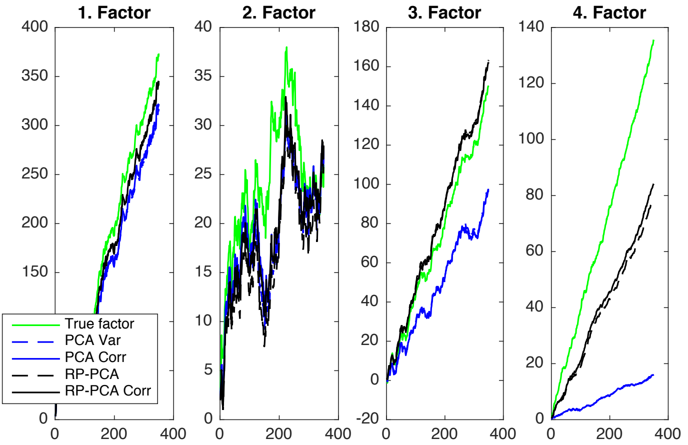


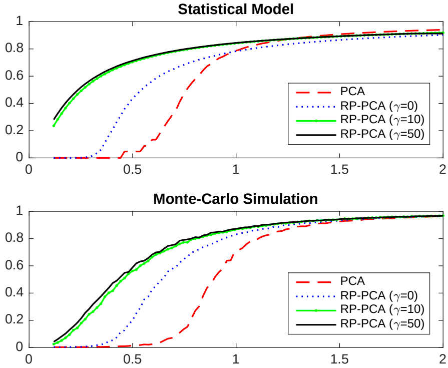
Figure: Sample path of the first four factors and the estimated factor processes. $\gamma = 50$.

Simulation

	PCA Var	PCA Corr	RP-PCA	RP-PCA Corr
1. Factor	0.094	0.086	0.042	0.040
2. Factor	0.023	0.022	0.025	0.022
3. Factor	0.100	0.095	0.079	0.074
4. Factor	**0.312**	**0.312**	**0.183**	**0.170**

Table: Average root-mean-squared (RMS) errors of estimated factors relative to the true factor processes. $\gamma = 50$.

Simulation



Squared correlations between estimated and true factor based on the weak factor model prediction and Monte-Carlo simulations for different variances of the factor. The Sharpe-ratio of the factor is 1, i.e. the mean equals $\mu_F = \sigma_F$. The normalized variance of the factors is $\sigma_F^2 \cdot N$.

Weak Factor Model: Dependent residuals

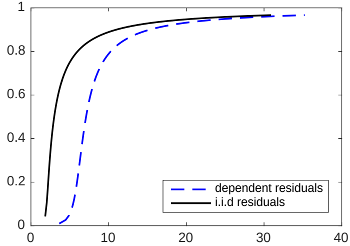


Figure: Values of ρ_i^2 ($\frac{1}{1+\theta_i B(\hat{\theta}_i))}$ if $\theta_i > \sigma_{crit}^2$ and 0 otherwise) for different signals θ_i. The average noise level is normalized in both cases to $\sigma_e^2 = 1$ and $c = 1$. For the correlated residuals we assume that $\Sigma^{1/2}$ is a Toeplitz matrix with $\beta, \beta, \beta, \beta^2$ on the right four off-diagonals with $\beta = 0.7$.

Simulation

	True Factors	PCA Var	PCA Corr	RP-PCA	PR-PCA Corr
SR	1.330	0.515	0.517	**0.865**	**0.883**

Table: Maximal Sharpe Ratio with $K = 4$ factors. $\gamma = 50$.

	True	PCA Var	PCA Corr	RP-PCA	RP-PCA Corr
1. Factor	1.20	1.10	1.11	1.16	1.16
2. Factor	0.10	0.11	0.10	0.12	0.11
3. Factor	0.40	0.31	0.31	0.49	0.48
4. Factor	**0.40**	**0.08**	**0.08**	**0.21**	**0.22**

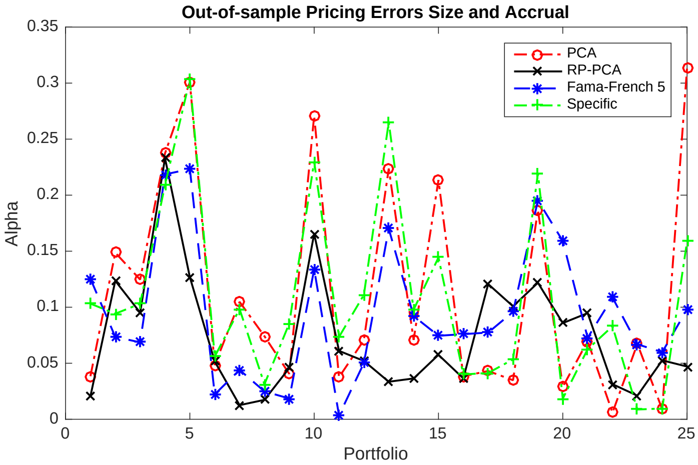
Table: Estimated mean of factors. $\gamma = 50$.

Simulation

	True	PCA Var	PCA Corr	RP-PCA	RP-PCA Corr
1. Factor	9.000	8.608	8.615	8.494	8.510
2. Factor	1.000	0.697	0.716	0.683	0.706
3. Factor	1.000	0.801	0.820	0.674	0.690
4. Factor	**0.160**	**0.028**	**0.028**	**0.066**	**0.070**

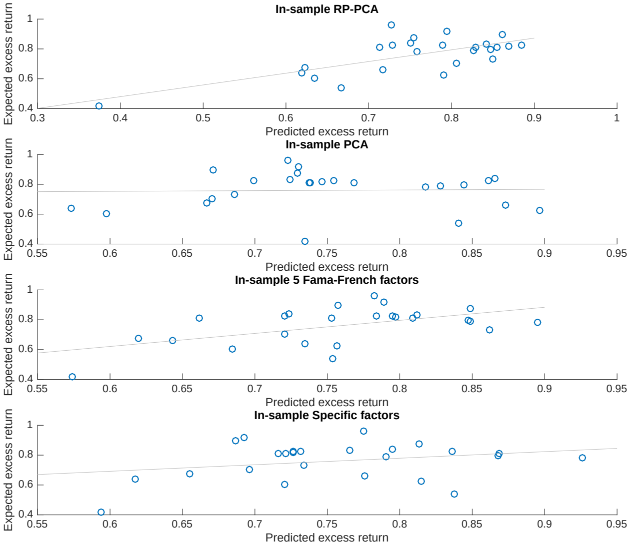
Table: Estimated variance of factors. $\gamma = 50$.

Cross-sectional α's out-of-sample (Size and Accrual)

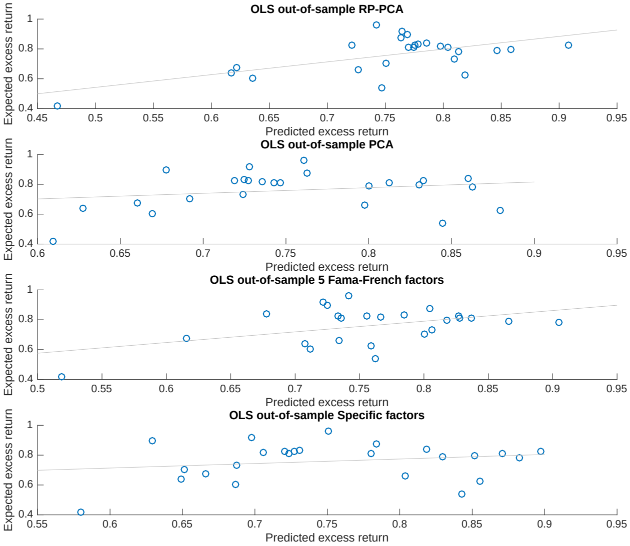


$\Rightarrow$ RP-PCA avoids large pricing errors due to penalty term.

Predicted excess return in-sample (Size and Accrual)



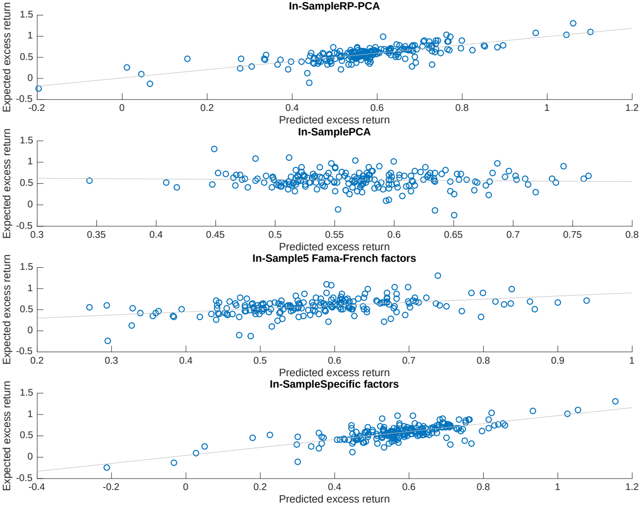
Predicted excess return out-of-sample (Size and Accrual)



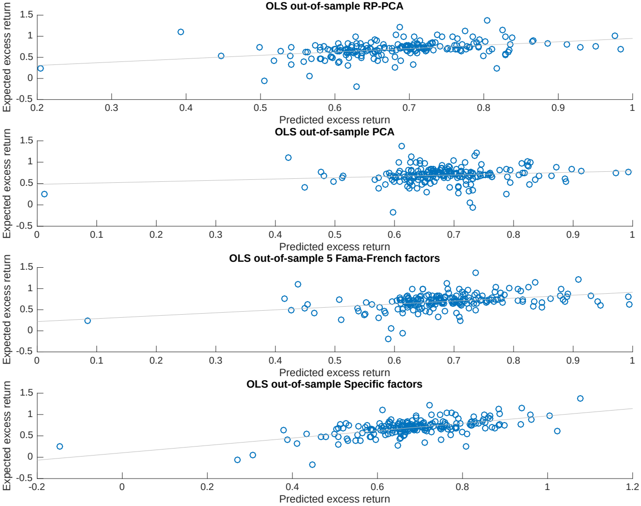
Fama-MacBeth Test-Statistics: χ^2_{22}: 34(95 %)

	RP-PCA	PCA	FF 5	Specific
Size and BM	85.66	94.50	79.99	105.15
BM and Investment	14.52	37.04	26.14	31.61
BM and Operating Profits	19.45	25.95	15.40	21.92
Size and Accrual	44.57	89.95	61.98	76.04
Size and Beta	30.74	32.90	31.76	31.96
Size and Investment	87.89	104.53	93.88	103.60
Size and Operating Profits	29.17	32.98	29.16	42.32
Size and Short-Term Reversal	87.70	103.35	88.86	108.31
Size and Long-Term Reversal	53.92	65.07	44.09	68.69
Size and Res. Var.	134.57	147.18	125.28	163.77
Size and Total Var.	120.14	133.46	120.71	143.01
Operating Profits and Investment	29.21	51.63	34.38	35.89
Size and Net Share Iss.	121.13	149.78	119.91	126.64
49 Industries	140.76	175.77	140.59	206.47

Predicted excess return in-sample



Predicted excess return out-of-sample



Maximal Incremental Sharpe Ratio

	PCA	RP-PCA
1 Factor	0.127	0.137
2 Factors	0.149	**0.381**
3 Factors	0.153	**0.412**

Table: Maximal Sharpe-ratio by adding factors incrementally. $K = 4$ statistical factors and risk-premium weight $\gamma = 100$.

Portfolio Data: Objective function

	PCA TS	RP-PCA TS	PCA XS	RP-PCA XS
1 Factor	44.771	51.623	0.298	0.037
2 Factors	39.846	42.326	0.268	0.001
3 Factors	36.112	37.849	0.263	0.000

Table: Time-series and cross-sectional objective functions.

$\Rightarrow$ RP-PCA and PCA explain the same amount of variation.

$\Rightarrow$ PR-PCA explains cross-sectional pricing much better.

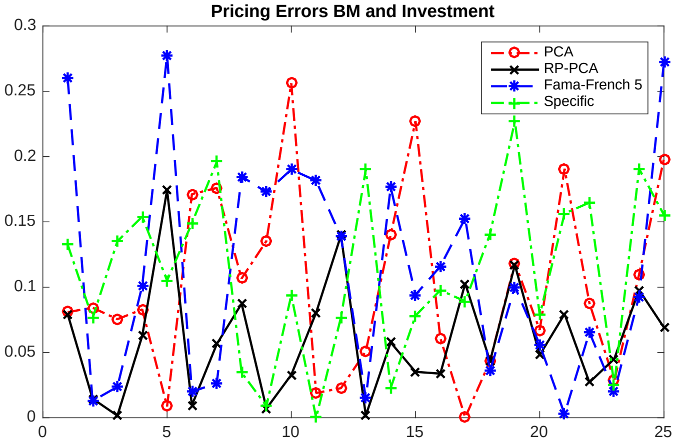
$\Rightarrow$ Motivation for risk-premium weight $\gamma = 100$.

Portfolio Data: In-sample (BM and Investment)

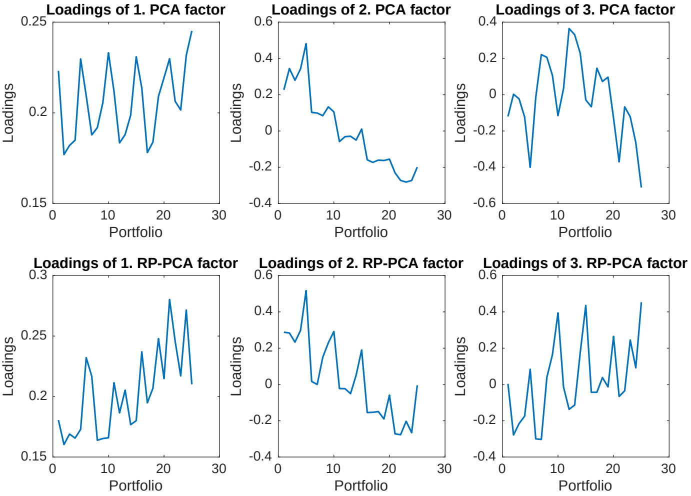
	SR	RMS α	Fama-MacBeth
RP-PCA	0.256	0.074	14.520
PCA	0.169	0.123	37.038
Fama-French	0.344	0.140	26.144
Specific	0.236	0.127	31.611

Table: Maximal Sharpe-ratios, root-mean-squared pricing errors and Fama-MacBeth test statistics for different set of factors. For the statistical factor estimators we use $K = 3$ factors and $\gamma = 100$.

Cross-sectional α's for sorted portfolios (BM and Investment)



Loadings for statistical factors (BM and Investment)



Maximal Incremental Sharpe Ratio (BM and Investment)

	PCA	RP-PCA
1 Factor	0.144	0.149
2 Factors	0.167	0.193
3 Factors	0.169	0.256

Table: Maximal Sharpe-ratio by adding factors incrementally. For the statistical factor estimators we use $K = 3$ factors and $\gamma = 100$.

Portfolio Data: Objective function (BM and Investment)

	PCA TS	RP-PCA TS	PCA XS	RP-PCA XS
1 Factor	5.543	5.989	0.021	0.002
2 Factors	4.416	4.647	0.014	0.001
3 Factors	3.944	4.098	0.013	0.000

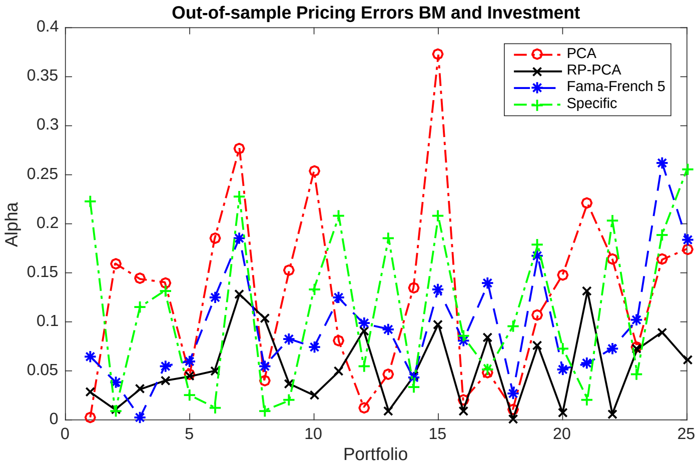
Table: Time-series and cross-sectional objective functions.

Portfolio Data: Out-of-sample (BM and Investment)

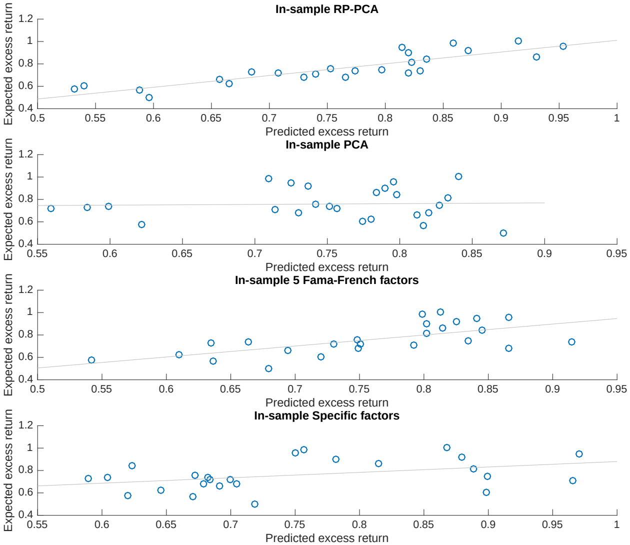
	Out-of-sample	In-sample
RP-PCA	0.123	0.065
PCA	0.157	0.156
Fama-French 5	0.111	0.103
Specific	0.138	0.138

Table: Root-mean-squared pricing errors for different set of factors. Out-of-sample factors are estimated with a rolling window. For the statistical factor estimators we use $K = 3$ factors and $\gamma = 100$.

Cross-sectional α's out-of-sample (BM and Investment)



Predicted excess return in-sample (BM and Investment)



Predicted excess return out-of-sample (BM and Invest.)