Factor Investing using Penalized Principal Components

Markus Pelger ¹  Martin Lettau ²

¹Stanford University
²UC Berkeley

February 8th 2018

AI in Fintech Forum 2018
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 300 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”

⇒ Summarize the pricing information of a large number of assets with a small number of factors
Why is it important?

Importance of factors for investing

1. Optimal portfolio construction
   - Only factors are compensated for systematic risk
   - Optimal portfolio with highest Sharpe-ratio must be based on factor portfolios
     (Sharpe-ratio = expected excess return/standard deviation)
   - “Smart beta” investments = exposure to risk factors

2. Arbitrage opportunities
   - Find underpriced assets and earn “alpha”

3. Risk management
   - Factors explain risk-return trade-off
   - Factors allow to manage systematic risk exposure
Contribution of this paper

**Contribution**

- This Paper: Estimation approach for finding risk factors and optimal factor investing
- 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:
  - Investment: 3 times higher Sharpe-ratio then benchmark factors (PCA)
  - Asset Pricing: Smaller pricing errors in- and out-of sample than benchmark (PCA, 5 Fama-French factors, etc.)
The Model

Approximate Factor Model

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

$$X_{t,i} = F_t^T \Lambda_i 1 \times K + e_{t,i} K \times 1$$

- Matrix notation

$$X_{T \times N} = F_{T \times K} \Lambda_{K \times N}^T + e_{T \times N}$$

- $N$ assets (large)
- $T$ time-series observation (large)
- $K$ systematic factors (fixed)

$F$, $\Lambda$ and $e$ are unknown
The Model

Approximate Factor Model

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

\[
\text{Var}(X) = \Lambda \text{Var}(F) \Lambda^\top + \text{Var}(e)
\]

\[
\underbrace{\text{systematic}}_{\text{systematic}} + \underbrace{\text{non-systematic}}_{\text{non-systematic}}
\]

⇒ 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:
  \[
  \frac{1}{T} X^T X - \bar{X}\bar{X}^T
  \]
  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
  \[
  \frac{1}{T} X^T X + \gamma \bar{X}\bar{X}^T \quad \gamma = \text{risk-premium weight}
  \]
- Eigenvectors of largest eigenvalues estimate loadings \( \hat{\Lambda} \).
- \( \hat{F} \) estimator for factors: \( \hat{F} = \frac{1}{N} X\hat{\Lambda} = X(\hat{\Lambda}^T\hat{\Lambda})^{-1}\hat{\Lambda}^T \).
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

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

where \( \bar{X}_i = \frac{1}{T} \sum_{t=1}^{T} X_{t,i} \) and \( \bar{F} = \frac{1}{T} \sum_{t=1}^{T} F_t \) and risk-premium weight \( \gamma \).
The Model: Interpretation

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:
   - 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

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

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

$$\Lambda (\Sigma_F + (1 + \gamma)\mu_F\mu_F^T) \Lambda^T + \text{Var}(e)$$

with $\Sigma_F = \text{Var}(F)$ and $\mu_F = E[F]$. The signal of weak factors with a small variance can be “pushed up” by their mean with the right $\gamma$. 
Illustration (Size and accrual)

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. **FF-long/short:** market, size and accrual (based on extreme quantiles, same construction as Fama-French factors)

- **Data**
  - Double-sorted portfolios according to size and accrual (from Kenneth French’s website)
  - Monthly return data from 07/1963 to 05/2017 ($T = 647$) for $N = 25$ portfolios
  - Out-of-sample: Rolling window of 20 years ($T=240$)

- Optimal factor investing: maximum Sharpe-ratio portfolio

\[
R_{opt} = F \cdot \Sigma_{F}^{-1} \mu_{F}
\]
### Illustration (Size and accrual)

<table>
<thead>
<tr>
<th></th>
<th>In-sample</th>
<th></th>
<th>Out-of-sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
<td>RMS $\alpha$</td>
<td>Idio. Var.</td>
<td>SR</td>
</tr>
<tr>
<td>RP-PCA</td>
<td>0.33</td>
<td>0.06</td>
<td>2.11</td>
<td>0.23</td>
</tr>
<tr>
<td>PCA</td>
<td>0.13</td>
<td>0.14</td>
<td>1.97</td>
<td>0.11</td>
</tr>
<tr>
<td>FF-long/short</td>
<td>0.21</td>
<td>0.12</td>
<td>2.63</td>
<td>0.11</td>
</tr>
</tbody>
</table>

**Table:** Maximal Sharpe-ratios, root-mean-squared pricing errors and unexplained idiosyncratic variation. $K = 3$ factors and $\gamma = 100$.

- **SR:** Maximum Sharpe-ratio of linear combination of factors
- **Cross-sectional pricing errors $\alpha$:**
  - Pricing error $\alpha_i = E[X_i] - E[F]\Lambda_i^T$
  - RMS $\alpha$: Root-mean-squared pricing errors $\sqrt{\frac{1}{N} \sum_{i=1}^{N} \alpha_i^2}$
  - Idiosyncratic Variation: $\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{t,i} - F_t^\top \Lambda_i)^2$

$\Rightarrow$ RP-PCA significantly better than PCA and quantile-sorted factors.
Loadings for statistical factors (Size and Accrual)

⇒ RP-PCA detects accrual factor while 3rd PCA factor is noise.
Maximal Sharpe ratio (Size and accrual)

![Graph showing maximal Sharpe ratio for different numbers of factors.](image)

**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 $\gamma > 10$. 

14
Weak vs. Strong Factors

The Model

Strong vs. weak factor models

- Strong factor model \((\frac{1}{N}\Lambda^T\Lambda \text{ bounded})\)
  - 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 \(\gamma\) relatively small

- Weak factor model \((\Lambda^T\Lambda \text{ bounded})\)
  - 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 \(\gamma\) relatively large

- In Data: Typically 1-2 strong factors and several weak factors with high Sharpe-ratios
Weak Factor Model

- Weak factors have small variance or affect small fraction of assets
- $\Lambda^T \Lambda$ bounded (after normalizing factor variances)
- Statistical model: Spiked covariance 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 distrib.
- Large eigenvalues converge either to
  - A biased value (characterized by the Stieltjes transform)
  - To the bulk of the spectrum if the true eigenvalue is too small

$\Rightarrow$ Phase transition phenomena: estimated eigenvectors orthogonal to true eigenvectors if eigenvalues too small
Intuition: Weak Factor Model

- “Signal” matrix for PCA of covariance matrix:
  \[ \Lambda \Sigma_F \Lambda^T \]
  
  The \( K \) largest eigenvalues \( \theta_1^{PCA}, ..., \theta_K^{PCA} \) measure the strength of the signal.

- “Signal” matrix for RP-PCA:
  \[ \Lambda (\Sigma_F + (1 + \gamma)\mu_F^T \mu_F) \Lambda^T \]
  
  The \( K \) largest eigenvalues \( \theta_1^{RP-PCA}, ..., \theta_K^{RP-PCA} \) also measure the strength of the signal.

- If \( \mu_F \neq 0 \) and \( \gamma > -1 \), then the RP-PCA signal is always larger than the PCA signal:
  \[ \theta_i^{RP-PCA} > \theta_i^{PCA} \]
Weak Factor Model

Theorem 1: Risk-Premium PCA under weak factor model

The correlation of the estimated with the true factors is

$$\hat{\text{Corr}}(F, \hat{F}) \overset{p}{\to} \begin{pmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \rho_K \end{pmatrix} \begin{pmatrix} \tilde{U} \\ \text{rotation} \end{pmatrix} \begin{pmatrix} \tilde{V} \\ \text{rotation} \end{pmatrix}$$

with

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

- Critical value $\theta_{\text{crit}}$ and function $B(.)$ depend only on the noise distribution and are known in closed-form.

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

- For $\theta_i > \theta_{\text{crit}}$, $\rho_i^2$ is strictly increasing in $\theta_i$.

$\Rightarrow$ RP-PCA strictly dominates PCA.
Strong Factor Model

Strong factors affect most assets: e.g. market factor

\[ \frac{1}{N} \Lambda^\top \Lambda \text{ bounded (after normalizing factor variances)} \]

Factors and loadings can be estimated consistently and are asymptotically normal distributed

Assumptions more general than in weak factor model

Asymptotic Efficiency

Choose RP-weight \( \gamma \) to obtain smallest asymptotic variance of estimators

- RP-PCA (i.e. \( \gamma > -1 \)) always more efficient than PCA
- In “most cases” optimal \( \gamma = 0 \) (smaller than in 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 $Z_{i,t-1,l}$:

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

- 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: Approximate non-linear function $g$ by basis functions
  
  ⇒ Project return data on appropriate basis functions
  
  ⇒ Equivalent to apply RP-PCA to portfolio data
  
  - Special case: Characteristics sorted portfolios corresponds to kernel basis functions
  
  - General case: Obtain arbitrary interactions and break curse of dimensionality by conditional tree sorting projection
Empirical Results

Single-sorted portfolios

Portfolio Data

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

<table>
<thead>
<tr>
<th></th>
<th>In-sample</th>
<th>Out-of-sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
<td>RMS</td>
</tr>
<tr>
<td>RP-PCA</td>
<td>0.66</td>
<td>0.15</td>
</tr>
<tr>
<td>PCA</td>
<td>0.28</td>
<td>0.15</td>
</tr>
<tr>
<td>Fama-French 5</td>
<td>0.32</td>
<td>0.23</td>
</tr>
</tbody>
</table>


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

Single-sorted portfolios: Maximal Sharpe-ratio

Figure: Maximal Sharpe-ratios.

⇒ Spike in Sharpe-ratio for 6 factors
Empirical Results

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
Single-sorted portfolios: Maximal Sharpe-ratio

**Figure:** Maximal Sharpe-ratios for different RP-weights $\gamma$ and number of factors $K$

$\Rightarrow$ Strong increase in Sharpe-ratios for $\gamma \geq 10$. 
Optimal Portfolio with RP-PCA

**Figure:** Portfolio composition of highest Sharpe ratio portfolio (Stochastic Discount Factor) based on 6 RP-PCA factors.
Optimal Portfolio with RP-PCA (largest positions)

**Figure**: Largest portfolios in highest Sharpe ratio portfolio (Stochastic Discount Factor) based on 6 RP-PCA factors.
Optimal Portfolio with PCA

Figure: Portfolio composition of highest Sharpe ratio portfolio (Stochastic Discount Factor) based on 6 PCA factors.
Optimal Portfolio with PCA (largest positions)

Figure: Largest portfolios in highest Sharpe ratio portfolio (Stochastic Discount Factor) based on 6 PCA factors.
Interpreting factors: Generalized correlations with proxies

<table>
<thead>
<tr>
<th></th>
<th>RP-PCA</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gen. Corr.</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2. Gen. Corr.</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>4. Gen. Corr.</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>5. Gen. Corr.</td>
<td>0.92</td>
<td>0.94</td>
</tr>
<tr>
<td>6. Gen. Corr.</td>
<td>0.78</td>
<td>0.89</td>
</tr>
</tbody>
</table>

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.
### Interpreting factors: 6th proxy factor

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Momentum (6m) 1</td>
<td>0.28</td>
<td>Leverage 10</td>
<td>0.33</td>
</tr>
<tr>
<td>Momentum (6m) 2</td>
<td>0.25</td>
<td>Asset Turnover 10</td>
<td>0.25</td>
</tr>
<tr>
<td>Value (M) 10</td>
<td>0.25</td>
<td>Value-Profitability 10</td>
<td>0.25</td>
</tr>
<tr>
<td>Value-Momentum 1</td>
<td>0.23</td>
<td>Profitability 10</td>
<td>0.22</td>
</tr>
<tr>
<td>Industry Momentum 1</td>
<td>0.20</td>
<td>Asset Turnover 9</td>
<td>0.22</td>
</tr>
<tr>
<td>Industry Reversals 9</td>
<td>0.19</td>
<td>Sales/Price 10</td>
<td>0.20</td>
</tr>
<tr>
<td>Industry Momentum 2</td>
<td>0.19</td>
<td>Sales/Price 9</td>
<td>0.18</td>
</tr>
<tr>
<td>Momentum (6m) 3</td>
<td>0.18</td>
<td>Size 10</td>
<td>0.17</td>
</tr>
<tr>
<td>Idiosyncratic Volatility 2</td>
<td>-0.18</td>
<td>Value-Momentum-Profitability 1</td>
<td>-0.19</td>
</tr>
<tr>
<td>Industry Mom. Reversals</td>
<td>-0.18</td>
<td>Profitability 2</td>
<td>-0.19</td>
</tr>
<tr>
<td>Value-Momentum 8</td>
<td>-0.20</td>
<td>Value-Profitability 1</td>
<td>-0.20</td>
</tr>
<tr>
<td>Momentum (6m) 10</td>
<td>-0.21</td>
<td>Profitability 4</td>
<td>-0.20</td>
</tr>
<tr>
<td>Value-Momentum 9</td>
<td>-0.23</td>
<td>Value-Profitability 2</td>
<td>-0.20</td>
</tr>
<tr>
<td>Value-Momentum 10</td>
<td>-0.23</td>
<td>Profitability 1</td>
<td>-0.23</td>
</tr>
<tr>
<td>Short-Term Reversals 1</td>
<td>-0.24</td>
<td>Idiosyncratic Volatility 1</td>
<td>-0.24</td>
</tr>
<tr>
<td>Industry-Momentum 10</td>
<td>-0.24</td>
<td>Profitability 3</td>
<td>-0.25</td>
</tr>
<tr>
<td>Industry Rel. Reversals 1</td>
<td>-0.28</td>
<td>Asset Turnover 2</td>
<td>-0.28</td>
</tr>
<tr>
<td>Idiosyncratic Volatility 1</td>
<td>-0.38</td>
<td>Asset Turnover 1</td>
<td>-0.35</td>
</tr>
</tbody>
</table>
Time-stability of loadings

Figure: Time-varying rotated loadings for the first six factors. Loadings are estimated on a rolling window with 240 months.
Time-stability of loadings of 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 $T = 500$)
- Out-of-sample performance collapses
- 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
Conclusion

Methodology

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

Empirical Results

- Strongly dominates PCA of the covariance matrix.
- Potential to provide benchmark factors for horse races.
- RP-PCA factor investing outperforms benchmark models
Time-stability: Generalized correlations

Figure: Generalized correlations between loadings estimated on the whole time horizon $T = 638$ and a rolling window with 240 months for $K = 6$ factors.
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.
**Effect of Risk-Premium Weight $\gamma$**

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

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

- Large-dimensional factor models with strong factors
  - Bai (2003): Distribution theory
  - 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
  - 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
Extreme deciles of single-sorted portfolios

Portfolio Data

- Monthly return data from 07/1963 to 12/2016 ($T = 638$) for $N = 64$ portfolios
- Kozak, Nagel and Santosh (2017) data: 370 decile portfolios sorted according to 37 anomalies

⇒ Here we take only the lowest and highest decile portfolio for each anomaly ($N = 64$).

Factors:

1. **RP-PCA**: $K = 6$ and $\gamma = 100$.
2. **PCA**: $K = 6$
3. **Fama-French 5**: The five factor model of Fama-French (market, size, value, investment and operating profitability, all from Kenneth French's website).
4. **Proxy factors**: RP-PCA and PCA factors approximated with 8 largest positions.
## Extreme Deciles

Table: Long-Short Portfolios of extreme deciles of 37 single-sorted portfolios from 07/1963 to 12/2016: Mean, standard deviation and Sharpe-ratio.

<table>
<thead>
<tr>
<th>Anomaly</th>
<th>Mean</th>
<th>SD</th>
<th>Sharpe-ratio</th>
<th>Anomaly</th>
<th>Mean</th>
<th>SD</th>
<th>Sharpe-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accruals</td>
<td>0.37</td>
<td>3.20</td>
<td>0.12</td>
<td>Momentum</td>
<td>1.28</td>
<td>6.91</td>
<td>0.19</td>
</tr>
<tr>
<td>Asset Turnover</td>
<td>0.40</td>
<td>3.84</td>
<td>0.10</td>
<td>Momentum-Reversals</td>
<td>0.47</td>
<td>4.82</td>
<td>0.10</td>
</tr>
<tr>
<td>Cash Flows/Price</td>
<td>0.44</td>
<td>4.38</td>
<td>0.10</td>
<td>Net Operating Assets</td>
<td>0.15</td>
<td>5.44</td>
<td>0.03</td>
</tr>
<tr>
<td>Composite Issuance</td>
<td>0.46</td>
<td>3.31</td>
<td>0.14</td>
<td>Price</td>
<td>0.03</td>
<td>6.82</td>
<td>0.00</td>
</tr>
<tr>
<td>Dividend/Price</td>
<td>0.2</td>
<td>5.11</td>
<td>0.04</td>
<td>Gross Profitability</td>
<td>0.36</td>
<td>3.41</td>
<td>0.11</td>
</tr>
<tr>
<td>Earnings/Price</td>
<td>0.57</td>
<td>4.76</td>
<td>0.12</td>
<td>Return on Assets (A)</td>
<td>0.21</td>
<td>4.07</td>
<td>0.05</td>
</tr>
<tr>
<td>Gross Margins</td>
<td>0.02</td>
<td>3.34</td>
<td>0.01</td>
<td>Return on Book Equity (A)</td>
<td>0.08</td>
<td>4.40</td>
<td>0.02</td>
</tr>
<tr>
<td>Asset Growth</td>
<td>0.33</td>
<td>3.46</td>
<td>0.10</td>
<td>Seasonality</td>
<td>0.81</td>
<td>3.94</td>
<td>0.21</td>
</tr>
<tr>
<td>Investment Growth</td>
<td>0.37</td>
<td>2.69</td>
<td>0.14</td>
<td>Sales Growth</td>
<td>0.05</td>
<td>3.59</td>
<td>0.01</td>
</tr>
<tr>
<td>Industry Momentum</td>
<td>0.49</td>
<td>6.17</td>
<td>0.08</td>
<td>Share Volume</td>
<td>0.00</td>
<td>6.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Industry Mom. Reversals</td>
<td>1.18</td>
<td>3.48</td>
<td>0.34</td>
<td>Size</td>
<td>0.29</td>
<td>4.81</td>
<td>0.06</td>
</tr>
<tr>
<td>Industry Rel. Reversals</td>
<td>1.00</td>
<td>4.11</td>
<td>0.24</td>
<td>Sales/Price</td>
<td>0.53</td>
<td>4.26</td>
<td>0.13</td>
</tr>
<tr>
<td>Industry Rel. Rev. (L.V.)</td>
<td>1.34</td>
<td>3.01</td>
<td>0.44</td>
<td>Short-Term Reversals</td>
<td>0.36</td>
<td>5.27</td>
<td>0.07</td>
</tr>
<tr>
<td>Investment/Assets</td>
<td>0.49</td>
<td>3.09</td>
<td>0.16</td>
<td>Value-Momentum</td>
<td>0.51</td>
<td>5.05</td>
<td>0.10</td>
</tr>
<tr>
<td>Investment/Capital</td>
<td>0.13</td>
<td>5.02</td>
<td>0.03</td>
<td>Value-Momentum-Prof.</td>
<td>0.84</td>
<td>4.85</td>
<td>0.17</td>
</tr>
<tr>
<td>Idiosyncratic Volatility</td>
<td>0.56</td>
<td>7.22</td>
<td>0.08</td>
<td>Value-Profitability</td>
<td>0.76</td>
<td>3.84</td>
<td>0.20</td>
</tr>
<tr>
<td>Leverage</td>
<td>0.24</td>
<td>4.58</td>
<td>0.05</td>
<td>Value (A)</td>
<td>0.50</td>
<td>4.57</td>
<td>0.11</td>
</tr>
<tr>
<td>Long Run Reversals</td>
<td>0.46</td>
<td>5.02</td>
<td>0.09</td>
<td>Value (M)</td>
<td>0.43</td>
<td>5.89</td>
<td>0.07</td>
</tr>
<tr>
<td>Momentum (6m)</td>
<td>0.35</td>
<td>6.27</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Extreme Deciles

<table>
<thead>
<tr>
<th></th>
<th>In-sample</th>
<th></th>
<th></th>
<th>Out-of-sample</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
<td>RMS $\alpha$</td>
<td>Idio. Var.</td>
<td>SR</td>
<td>RMS $\alpha$</td>
<td>Idio. Var.</td>
</tr>
<tr>
<td>RP-PCA</td>
<td>0.64</td>
<td>0.18</td>
<td>3.59</td>
<td>0.53</td>
<td>0.15</td>
<td>4.23</td>
</tr>
<tr>
<td>PCA</td>
<td>0.35</td>
<td>0.22</td>
<td>3.57</td>
<td>0.28</td>
<td>0.19</td>
<td>4.24</td>
</tr>
<tr>
<td>RP-PCA Proxy</td>
<td>0.62</td>
<td>0.19</td>
<td>4.08</td>
<td>0.48</td>
<td>0.17</td>
<td>4.19</td>
</tr>
<tr>
<td>PCA Proxy</td>
<td>0.37</td>
<td>0.22</td>
<td>3.77</td>
<td>0.315</td>
<td>0.18</td>
<td>4.20</td>
</tr>
<tr>
<td>Fama-French 5</td>
<td>0.32</td>
<td>0.30</td>
<td>7.31</td>
<td>0.31</td>
<td>0.262</td>
<td>6.40</td>
</tr>
</tbody>
</table>


- RP-PCA strongly dominates PCA and Fama-French 5 factors
- Results hold out-of-sample.
Extreme Deciles: Maximal Sharpe-ratio

Figure: Maximal Sharpe-ratios.

⇒ Spike in Sharpe-ratio for 6 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
Figure: First and last decile of 37 single-sorted portfolios ($N = 64$ and $T = 638$): Maximal Sharpe-ratios for different RP-weights $\gamma$ and number of factors $K$. 
Interpreting factors: Generalized correlations with proxies

<table>
<thead>
<tr>
<th></th>
<th>RP-PCA</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gen. Corr.</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2. Gen. Corr.</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3. Gen. Corr.</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>4. Gen. Corr.</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>5. Gen. Corr.</td>
<td>0.88</td>
<td>0.95</td>
</tr>
<tr>
<td>6. Gen. Corr.</td>
<td>0.72</td>
<td>0.89</td>
</tr>
</tbody>
</table>

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.
## Interpreting factors: Cumulative absolute proxy weights

<table>
<thead>
<tr>
<th>RP-PCA Proxy</th>
<th>PCA Proxy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momentum (12m)</td>
<td>Idiosyncratic Volatility</td>
</tr>
<tr>
<td>1.70</td>
<td>1.28</td>
</tr>
<tr>
<td>Idiosyncratic Volatility</td>
<td>Momentum (12m)</td>
</tr>
<tr>
<td>1.65</td>
<td>1.14</td>
</tr>
<tr>
<td>Industry Rel. Rev. (L.V.)</td>
<td>Momentum (6m)</td>
</tr>
<tr>
<td>1.21</td>
<td>1.04</td>
</tr>
<tr>
<td>Momentum (6m)</td>
<td>Price</td>
</tr>
<tr>
<td>1.21</td>
<td>0.95</td>
</tr>
<tr>
<td>Price</td>
<td></td>
</tr>
<tr>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>Industry Mom. Reversals</td>
<td>Asset Turnover</td>
</tr>
<tr>
<td>1.11</td>
<td>0.83</td>
</tr>
<tr>
<td>Value (M)</td>
<td>Industry Rel. Rev. (L.V.)</td>
</tr>
<tr>
<td>1.03</td>
<td>0.79</td>
</tr>
<tr>
<td>Industry Rel. Reversals</td>
<td>Value (M)</td>
</tr>
<tr>
<td>0.84</td>
<td>0.75</td>
</tr>
<tr>
<td>Share Volume</td>
<td>Industry Momentum</td>
</tr>
<tr>
<td>0.55</td>
<td>0.73</td>
</tr>
<tr>
<td>Net Operating Assets</td>
<td>Industry Mom. Reversals</td>
</tr>
<tr>
<td>0.47</td>
<td>0.73</td>
</tr>
<tr>
<td>Value-Momentum-Prof.</td>
<td>Dividend/Price</td>
</tr>
<tr>
<td>0.46</td>
<td>0.67</td>
</tr>
<tr>
<td>Size</td>
<td>Sales/Price</td>
</tr>
<tr>
<td>0.42</td>
<td>0.58</td>
</tr>
<tr>
<td>Return on Book Equity (A)</td>
<td>Value-Profitability</td>
</tr>
<tr>
<td>0.32</td>
<td>0.58</td>
</tr>
<tr>
<td>Industry Momentum</td>
<td>Net Operating Assets</td>
</tr>
<tr>
<td>0.29</td>
<td>0.37</td>
</tr>
<tr>
<td>Value-Momentum</td>
<td>Value (A)</td>
</tr>
<tr>
<td>0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>Long Run Reversals</td>
<td>Value-Momentum-Prof.</td>
</tr>
<tr>
<td>0.24</td>
<td>0.33</td>
</tr>
<tr>
<td>Earnings/Price</td>
<td>Cash Flows/Price</td>
</tr>
<tr>
<td>0.22</td>
<td>0.31</td>
</tr>
</tbody>
</table>
## Interpreting factors: Composition of proxies

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>indrrevlv10</td>
<td>0.54</td>
<td>valmomprof10</td>
<td>0.17</td>
<td>mom1210</td>
</tr>
<tr>
<td>indmomrev10</td>
<td>0.52</td>
<td>indmomrev10</td>
<td>-0.20</td>
<td>mom10</td>
</tr>
<tr>
<td>ivol1</td>
<td>0.24</td>
<td>ivol1</td>
<td>-0.21</td>
<td>valuem1</td>
</tr>
<tr>
<td>Accrual1</td>
<td>-0.21</td>
<td>mom1211</td>
<td>-0.23</td>
<td>lrrev10</td>
</tr>
<tr>
<td>shvol1</td>
<td>-0.22</td>
<td>indrrevlv10</td>
<td>-0.26</td>
<td>mom1</td>
</tr>
<tr>
<td>ep1</td>
<td>-0.22</td>
<td>indmomrev1</td>
<td>-0.40</td>
<td>valuem1</td>
</tr>
<tr>
<td>indrrevlv1</td>
<td>-0.25</td>
<td>indrrevlv1</td>
<td>-0.41</td>
<td>price1</td>
</tr>
<tr>
<td>mom1211</td>
<td>-0.42</td>
<td>ivol1</td>
<td>-0.67</td>
<td>mom1211</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ivol1</td>
<td>0.59</td>
<td>valuem10</td>
<td>0.46</td>
<td>mom10</td>
</tr>
<tr>
<td>indrrevlv10</td>
<td>0.43</td>
<td>price1</td>
<td>0.38</td>
<td>indmom10</td>
</tr>
<tr>
<td>indmomrev10</td>
<td>0.37</td>
<td>divp10</td>
<td>0.37</td>
<td>mom1210</td>
</tr>
<tr>
<td>indrrevlv1</td>
<td>0.36</td>
<td>value1</td>
<td>0.36</td>
<td>valmomprof10</td>
</tr>
<tr>
<td>indmomrev1</td>
<td>0.36</td>
<td>sp10</td>
<td>0.32</td>
<td>valmom1</td>
</tr>
<tr>
<td>ivol10</td>
<td>0.27</td>
<td>lrrev10</td>
<td>0.31</td>
<td>indmom1</td>
</tr>
<tr>
<td>mom121</td>
<td>0.03</td>
<td>cfp10</td>
<td>0.31</td>
<td>mom1211</td>
</tr>
<tr>
<td>indmom1</td>
<td>0.03</td>
<td>valuem1</td>
<td>-0.29</td>
<td>mom1</td>
</tr>
</tbody>
</table>

### 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

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

Figure: Generalized correlations between loadings estimated on the whole time horizon $T = 638$ and a rolling window.
Double-sorted portfolios

Data
- Monthly return data from 07/1963 to 05/2017 ($T = 647$)
- 13 double sorted portfolios (consisting of 25 portfolios) from Kenneth French's website

Factors
1. **PCA**: $K = 3$
2. **RP-PCA**: $K = 3$ and $\gamma = 100$
3. **FF-Long/Short** factors: market + two specific anomaly long-short factors
### Sharpe-ratios and pricing errors (in-sample)

<table>
<thead>
<tr>
<th></th>
<th>Sharpe-Ratio</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RPCA</td>
<td>PCA</td>
<td>FF-L/S</td>
<td>RPCA</td>
<td>PCA</td>
<td>FF-L/S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and BM</td>
<td>0.25</td>
<td>0.22</td>
<td>0.21</td>
<td>0.13</td>
<td>0.13</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM and Investment</td>
<td><strong>0.24</strong></td>
<td><strong>0.17</strong></td>
<td><strong>0.24</strong></td>
<td><strong>0.08</strong></td>
<td><strong>0.11</strong></td>
<td><strong>0.12</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM and Profits</td>
<td>0.23</td>
<td>0.20</td>
<td>0.24</td>
<td>0.10</td>
<td>0.12</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and Accrual</td>
<td><strong>0.33</strong></td>
<td><strong>0.13</strong></td>
<td><strong>0.21</strong></td>
<td><strong>0.06</strong></td>
<td><strong>0.14</strong></td>
<td><strong>0.12</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and Beta</td>
<td>0.25</td>
<td>0.24</td>
<td>0.23</td>
<td>0.06</td>
<td>0.07</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and Investment</td>
<td>0.32</td>
<td>0.26</td>
<td>0.21</td>
<td>0.10</td>
<td>0.11</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and Profits</td>
<td>0.22</td>
<td>0.21</td>
<td>0.25</td>
<td>0.06</td>
<td>0.06</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and Momentum</td>
<td>0.25</td>
<td>0.19</td>
<td>0.18</td>
<td>0.15</td>
<td>0.16</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and ST-Reversal</td>
<td>0.28</td>
<td>0.25</td>
<td>0.24</td>
<td>0.16</td>
<td>0.17</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and Idio. Vol.</td>
<td>0.35</td>
<td>0.31</td>
<td>0.32</td>
<td>0.15</td>
<td>0.16</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and Total Vol.</td>
<td>0.34</td>
<td>0.30</td>
<td>0.31</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profits and Invest.</td>
<td>0.28</td>
<td>0.24</td>
<td>0.30</td>
<td>0.10</td>
<td>0.12</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size and LT-Reversal</td>
<td>0.22</td>
<td>0.18</td>
<td>0.16</td>
<td>0.11</td>
<td>0.13</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sharpe-ratios and pricing errors (out-of-sample)

<table>
<thead>
<tr>
<th>Sharpe-Ratio</th>
<th>RPCA</th>
<th>PCA</th>
<th>FF-L/S</th>
<th>RPCA</th>
<th>PCA</th>
<th>FF-L/S</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size and BM</strong></td>
<td>0.23</td>
<td>0.18</td>
<td>0.16</td>
<td>0.18</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td><strong>BM and Investment</strong></td>
<td><strong>0.21</strong></td>
<td><strong>0.15</strong></td>
<td><strong>0.24</strong></td>
<td><strong>0.13</strong></td>
<td><strong>0.17</strong></td>
<td><strong>0.17</strong></td>
</tr>
<tr>
<td>BM and Profits</td>
<td>0.19</td>
<td>0.17</td>
<td>0.23</td>
<td>0.16</td>
<td>0.17</td>
<td>0.19</td>
</tr>
<tr>
<td><strong>Size and Accrual</strong></td>
<td><strong>0.24</strong></td>
<td><strong>0.09</strong></td>
<td><strong>0.11</strong></td>
<td><strong>0.08</strong></td>
<td><strong>0.14</strong></td>
<td><strong>0.12</strong></td>
</tr>
<tr>
<td>Size and Beta</td>
<td>0.22</td>
<td>0.20</td>
<td>0.16</td>
<td>0.08</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Size and Investment</td>
<td>0.30</td>
<td>0.23</td>
<td>0.17</td>
<td>0.13</td>
<td>0.14</td>
<td>0.16</td>
</tr>
<tr>
<td>Size and Profits</td>
<td>0.22</td>
<td>0.20</td>
<td>0.20</td>
<td>0.10</td>
<td>0.1</td>
<td>0.17</td>
</tr>
<tr>
<td>Size and Momentum</td>
<td>0.18</td>
<td>0.12</td>
<td>0.08</td>
<td>0.17</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>Size and ST-Reversal</td>
<td>0.22</td>
<td>0.17</td>
<td>0.23</td>
<td>0.22</td>
<td>0.23</td>
<td>0.25</td>
</tr>
<tr>
<td>Size and Idio. Vol.</td>
<td>0.37</td>
<td>0.30</td>
<td>0.28</td>
<td>0.17</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>Size and Total Vol.</td>
<td>0.35</td>
<td>0.28</td>
<td>0.27</td>
<td>0.18</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>Profits and Invest.</td>
<td>0.31</td>
<td>0.25</td>
<td>0.29</td>
<td>0.12</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>Size and LT-Reversal</td>
<td>0.16</td>
<td>0.10</td>
<td>0.04</td>
<td>0.13</td>
<td>0.14</td>
<td>0.14</td>
</tr>
</tbody>
</table>
The Model: Objective function

Time-series objective function:

Minimize the unexplained variance:

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

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

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

$\Rightarrow$ Motivation for PCA
The Model: Objective function

Cross-sectional objective function:

Minimize cross-sectional expected pricing error:

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \hat{E}[X_i] - \hat{E}[F] \Lambda_i^T \right)^2
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} X_i^T \mathbb{1} - \frac{1}{T} \mathbb{1}^T F \Lambda_i^T \right)^2
\]

\[
= \frac{1}{N} \text{trace} \left( \left( \frac{1}{T} \mathbb{1}^T X \Lambda \right) \left( \frac{1}{T} \mathbb{1}^T X \Lambda \right)^T \right) \quad \text{s.t. } F = X (\Lambda^T \Lambda)^{-1} \Lambda^T
\]

- \( \mathbb{1} \) is vector \( T \times 1 \) of 1’s and thus \( F^T \mathbb{1} / T \) 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

\[
\min_{\Lambda, F} \frac{1}{NT} \text{trace} \left( \left( (XM_\Lambda)^\top (XM_\Lambda) \right) \right) + \gamma \frac{1}{N} \text{trace} \left( \left( \frac{1}{T} 1_1^\top X M_\Lambda \right) \left( \frac{1}{T} 1_1^\top X M_\Lambda \right)^\top \right)
\]

\[
= \min_{\Lambda} \frac{1}{NT} \text{trace} \left( M_\Lambda X^\top \left( I + \frac{\gamma}{T} 1_1 1_1^\top \right) X M_\Lambda \right) \quad \text{s.t. } F = X (\Lambda^\top \Lambda)^{-1} \Lambda^\top
\]

- The objective function is minimized by the eigenvectors of the largest eigenvalues of \( \frac{1}{NT} X^\top \left( I_T + \frac{\gamma}{T} 1_1 1_1^\top \right) X \).
- \( \hat{\Lambda} \) estimator for loadings: proportional to eigenvectors of the first \( K \) eigenvalues of \( \frac{1}{NT} X^\top \left( I_T + \frac{\gamma}{T} 1_1 1_1^\top \right) X \)
- \( \hat{F} \) 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:

\[
\min_{\Lambda, F} \frac{1}{NT} \text{trace}(Q^T (X - F\Lambda^T)^T (X - F\Lambda^T)Q) + \gamma \frac{1}{N} \text{trace} \left( \mathbf{1}^T (X - F\Lambda^T)QQ^T (X - F\Lambda^T)^T \mathbf{1} \right)
\]

\[
= \min_{\Lambda} \text{trace} \left( M_{\Lambda} Q^T X^T \left( I + T \frac{T}{T} \right) XQM_{\Lambda} \right) \quad \text{s.t.} \quad F = X(\Lambda^T \Lambda)^{-1} \Lambda^T
\]

- Cross-sectional weighting matrix \( Q \)
- Factors and loadings can be estimated by applying PCA to \( Q^T X^T \left( I + \frac{T}{T} \mathbf{1} \mathbf{1}^T \right) XQ \).
- 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
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 $\frac{1}{N} \Lambda^\top \Lambda = I_K$.
  $\Lambda_{i, 1} = 1$ 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. 1/3 of variation due to systematic risk.
Simulation parameters

Errors

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

- $\epsilon$ 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 $(\beta, \beta, \beta)$ on the right three 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: $\sigma_e^2 = 9$
- Parameters produce eigenvalues that are consistent with the data.
Simulation

Figure: Sample paths of the cumulative returns of the first four factors and the estimated factor processes. $N = 250$ and $T = 350$. 
## Simulation

<table>
<thead>
<tr>
<th>True Factor</th>
<th>$\gamma = -1$</th>
<th>$\gamma = 0$</th>
<th>$\gamma = 10$</th>
<th>$\gamma = 50$</th>
<th>$\gamma = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In-sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>1.17</td>
<td>0.56</td>
<td>0.80</td>
<td>1.07</td>
<td>1.10</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.42</td>
<td>0.57</td>
<td>0.51</td>
<td>0.42</td>
<td>0.41</td>
</tr>
<tr>
<td>Idio. Var.</td>
<td>23.77</td>
<td>22.728</td>
<td>22.784</td>
<td>22.912</td>
<td>22.923</td>
</tr>
<tr>
<td>Ex. Var.</td>
<td>0.33</td>
<td>0.36</td>
<td>0.36</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td><strong>Out-of-sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>1.17</td>
<td>0.60</td>
<td>0.79</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.42</td>
<td>0.58</td>
<td>0.52</td>
<td>0.47</td>
<td>0.47</td>
</tr>
<tr>
<td>Idio. Var.</td>
<td>23.84</td>
<td>23.15</td>
<td>23.13</td>
<td>23.15</td>
<td>23.15</td>
</tr>
<tr>
<td>Ex. Var.</td>
<td>0.33</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
</tbody>
</table>

**Table:** (i) Maximal Sharpe-ratio (SR), (ii) root-mean-squared pricing error ($\alpha$), unexplained variation and proportion of variation explained by $K = 4$ factors. $N = 250$ and $T = 350$. 

$\gamma = -1, 0, 10, 50, 100$.
Simulation

<table>
<thead>
<tr>
<th></th>
<th>$\gamma = -1$</th>
<th>$\gamma = 0$</th>
<th>$\gamma = 10$</th>
<th>$\gamma = 50$</th>
<th>$\gamma = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In-sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Factor</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>2. Factor</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>3. Factor</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>4. Factor</td>
<td><strong>0.15</strong></td>
<td><strong>0.45</strong></td>
<td><strong>0.68</strong></td>
<td><strong>0.69</strong></td>
<td><strong>0.69</strong></td>
</tr>
<tr>
<td><strong>Out-of-sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Factor</td>
<td>0.94</td>
<td>0.97</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>2. Factor</td>
<td>0.79</td>
<td>0.88</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>3. Factor</td>
<td>0.77</td>
<td>0.86</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>4. Factor</td>
<td><strong>0.17</strong></td>
<td><strong>0.46</strong></td>
<td><strong>0.68</strong></td>
<td><strong>0.69</strong></td>
<td><strong>0.69</strong></td>
</tr>
</tbody>
</table>

Table: Correlation between estimated and true factors. The risk-premium weight $\gamma = -1$ corresponds to PCA. $N = 250$ and $T = 350$. 
Simulation

Simulation parameters

- Consider only $K = 1$ factor
- Sharpe-ratio $= 1$, i.e. $\mu_F = \sigma_F$
- Different signal strength $\sigma_F^2$
- $\Lambda_i \sim N(0, 1)$, i.e. normalized variance $\sigma_F^2 \cdot N$ in weak factor model.
- Same residuals as before
Simulation: Correlation, N=150, T=150

Figure: Correlation of estimated with true factor.

⇒ For signal $\sigma^2 < 0.8$ a value of $\gamma \geq 10$ is optimal.
Simulation: Sharpe-ratio, N=150, T=150

Figure: Maximal Sharpe-ratios.

⇒ For signal $\sigma^2 < 0.8$ a value of $\gamma \geq 10$ is optimal.
Simulation: Pricing Error, N=150, T=150

Figure: Root-mean-squared pricing errors.

⇒ For signal $\sigma^2 < 0.8$ a value of $\gamma \geq 10$ is optimal.
Simulation: Idiosyncratic Variation, N=150, T=150

Figure: Unexplained idiosyncratic variation.

⇒ Unexplained variation similar for RP-PCA and PCA.
Simulation: Correlation, N=250, T=250

Figure: Correlation of estimated with true factor.

⇒ For signal $\sigma^2 < 0.8$ a value of $\gamma \geq 0$ is optimal.
Simulation: Correlation, N=50, T=250

Figure: Correlation of estimated with true factor.

⇒ For signal $\sigma^2 < 0.8$ a value of $\gamma \geq 10$ is optimal.
Simulation: Correlation, N=100, T=250

Figure: Correlation of estimated with true factor.

⇒ For signal $\sigma^2 < 0.8$ a value of $\gamma \geq 10$ is optimal.
Simulation: Correlation, N=250, T=150

Figure: Correlation of estimated with true factor.

⇒ For signal $\sigma^2 < 0.8$ a value of $\gamma \geq 0$ is optimal.
Simulation: Weak factor model prediction

Correlations between estimated and true factor based on the weak factor model prediction and Monte-Carlo simulations for different variances of the factor. The residuals have cross-sectional and time-series correlation and heteroskedasticity. The Sharpe-ratio of the factor is 1. \( N = 250, T = 350. \)
Correlations between estimated and true factor based on the weak factor model prediction and Monte-Carlo simulations for different variances of the factor. The residuals have only cross-sectional correlation. The Sharpe-ratio of the factor is 1. $N = 250$, $T = 350$. 
Weak Factor Model: Dependent residuals

**Figure:** Values of $\rho_i^2 \left( \frac{1}{1+\theta_i B(\hat{\theta}_i)} \right)$ if $\theta_i > \sigma^2_{\text{crit}}$ and 0 otherwise) for different signals $\theta_i$. The average noise level is normalized in both cases to $\sigma^2_e = 1$ and $c = 1$. For the correlated residuals we assume that $\Sigma^{1/2}$ is a Toeplitz matrix with $\beta, \beta, \beta$ on the right three off-diagonals with $\beta = 0.7$. $N = 250$, $T = 350$. 

\[\text{Figure: } \rho_i^2 \left( \frac{1}{1+\theta_i B(\hat{\theta}_i)} \right) \text{ if } \theta_i > \sigma^2_{\text{crit}} \text{ and 0 otherwise.} \]
Weak Factor Model

- Weak factors either have a small variance or affect a smaller fraction of assets:
- \( \Lambda^T \Lambda \) 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

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 \textbf{mean zero}
  ⇒ RP-PCA implicitly uses non-zero means of random variables
  ⇒ New tools necessary!
Assumption 1: Weak Factor Model

1. **Rate:** Assume that $\frac{N}{T} \rightarrow c$ with $0 < c < \infty$.

2. **Factors:** $F$ are uncorrelated among each other and are independent of $e$ and $\Lambda$ and have bounded first two moments.

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

3. **Loadings:** The column vectors of the loadings $\Lambda$ are orthogonally invariant and independent of $\epsilon$ 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 $\Sigma$ converges to a non-random spectral distribution function with compact support and supremum of support $b$. Largest eigenvalues of $\Sigma$ converge to $b$. 
Weak Factor Model

Intuition: Weak Factor Model

- “Signal” matrix for PCA of covariance matrix:
  \[ \Lambda \Sigma_F \Lambda^T \]
  
  \( K \) largest eigenvalues \( \theta_{1}^{PCA}, ..., \theta_{K}^{PCA} \) measure strength of signal

- “Signal” matrix for RP-PCA:
  \[ \Lambda \left( \Sigma_F + (1 + \gamma)\mu_F \mu_F^T \right) \Lambda^T \]
  
  \( K \) largest eigenvalues \( \theta_{1}^{RP-PCA}, ..., \theta_{K}^{RP-PCA} \) measure strength of signal

- If \( \mu_F \neq 0 \) and \( \gamma > -1 \) then RP-PCA signal always larger than PCA signal:
  \[ \theta_{i}^{RP-PCA} > \theta_{i}^{PCA} \]
**Theorem 1: Risk-Premium PCA under weak factor model**

Under Assumption 1 the correlation of the estimated with the true factors is

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

with

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

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

\(\Rightarrow\) RP-PCA strictly dominates PCA.
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 $\Lambda$ 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 $W^2 = \left( I_T + \gamma \frac{11^\top}{T} \right)$ and $1$ is a $T \times 1$ vector of 1's.
  - Asymptotically $\hat{F}$ behaves like OLS regression of $\Lambda$ on $X$.

Asymptotic Efficiency

Choose RP-weight $\gamma$ to obtain smallest asymptotic variance of estimators

- RP-PCA (i.e. $\gamma > -1$) always more efficient than PCA
- Optimal $\gamma$ 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:** \( \Lambda^T \Lambda / N \xrightarrow{P} I_K \) 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 \( \Sigma \) are bounded.
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 $\gamma$.
3. The asymptotic distribution of the loadings is given by
   \[
   \sqrt{T} \left( \hat{\Lambda}_i - \Lambda_i \right) \xrightarrow{D} N(0, \Omega_i)
   \]
   \[
   \Omega_i = \sigma_{e_i}^2 \left( \Sigma_F + (1 + \gamma) \mu_F \mu_F^\top \right)^{-1} \left( \Sigma_F + (1 + \gamma)^2 \mu_F \mu_F^\top \right)
   \]
   \[
   \left( \Sigma_F + (1 + \gamma) \mu_F \mu_F^\top \right)^{-1}, \quad E[e_{t,i}^2] = \sigma_{e_i}^2
   \]
4. $\gamma = 0$ is **optimal choice** for smallest asymptotic variance.
   $\gamma = -1$, i.e. the covariance matrix, is not efficient.
Model with time-varying loadings

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

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

- 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 $\phi_m(.)$:

$$g_k(Z_{i,t-1}) = \sum_{m=1}^{M} b_{m,k} \phi_m(Z_{i,t-1})$$

$$g(Z_{t-1}) = B^\top \Phi(Z_{t-1})$$

- 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: $\phi_m = 1_{\{Z_{t-1} \in I_m\}} \Rightarrow \tilde{X}$ 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 $Z_{t-1}$. 
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 $\Sigma$ 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 $\Lambda$ and have bounded first two moments.

$$\hat{\mu}_F := \frac{1}{T} \sum_{t=1}^{T} F_t \xrightarrow{p} \mu_F$$

$$\hat{\Sigma}_F := \frac{1}{T} F_t F_t^\top \xrightarrow{p} \Sigma_F = \begin{pmatrix} \sigma_{F_1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{F_K}^2 \end{pmatrix}$$

3. The column vectors of the loadings $\Lambda$ are orthogonally invariant and independent of $\epsilon$ 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 $0 < c < \infty$. 
Definition: Weak Factor Model

- Average idiosyncratic noise $\sigma_e^2 := \text{trace}(\Sigma)/N$
- Denote by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ 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:

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

- $B$-function

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

$$= \lim_{T \to \infty} \frac{c}{N} \text{trace} \left( \left( (zI_N - \frac{1}{T}e^\top e) \right)^{-2} \left( \frac{1}{T}e^\top e \right) \right)$$
Weak Factor Model

**Estimator**

- Risk-premium PCA (RP-PCA): Apply PCA estimation to
  \[ S_\gamma := \frac{1}{T}X^\top \left( I_T + \gamma \frac{1 1^\top}{T} \right) X \]

- PCA: Apply PCA to estimated covariance matrix
  \[ S_{-1} := \frac{1}{T}X^\top \left( I_T - \frac{1 1^\top}{T} \right) X, \text{ i.e. } \gamma = -1. \]

⇒ PCA special case of RP-PCA

**“Signal” Matrix for Covariance PCA**

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

⇒ Intuition: Largest \( K \) “true” eigenvalues of \( S_{-1} \).
Lemma: Covariance PCA

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

$$
\hat{\lambda}_i \overset{p}{\rightarrow} \begin{cases} 
G^{-1} \left( \frac{1}{\sigma_{F_i}^2 + c\sigma_e^2} \right) & \text{if } \sigma_{F_i}^2 + c\sigma_e^2 > \sigma_{\text{crit}}^2 \\
 b & \text{otherwise}
\end{cases}
$$

The correlation between the estimated and true factors converges to

$$
\widehat{\text{Corr}}(F, \hat{F}) \overset{p}{\rightarrow} \begin{pmatrix} \rho_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \rho_K \end{pmatrix}
$$

with

$$
\rho_i^2 \overset{p}{\rightarrow} \begin{cases} 
\frac{1}{1 + (\sigma_{F_i}^2 + c\sigma_e^2)B(\hat{\lambda}_i)} & \text{if } \sigma_{F_i}^2 + c\sigma_e^2 > \sigma_{\text{crit}}^2 \\
0 & \text{otherwise}
\end{cases}
$$
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:

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

The correlation between the estimated and true factors converges to

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

with

$$
Q^2_i \xrightarrow{p} \begin{cases} 
1 - \frac{c\sigma^4_e}{\sigma^4_F} & \text{if } \sigma^2_F + c\sigma^2_e > \sigma^2_{\text{crit}} \\
\frac{1 + \frac{\sigma^4_e}{\sigma^4_F} (c^2 - c)}{\frac{\sigma^4_e}{\sigma^4_F} + \frac{\sigma^4_e}{\sigma^4_F} (c^2 - c)} & \text{otherwise}
\end{cases}
$$
Weak Factor Model

“Signal” Matrix for RP-PCA

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

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

⇒ Projection on \( K \) demeaned factors and on mean operator.

Denote by \( \theta_1 \geq \ldots \geq \theta_{K+1} \) the eigenvalues of the “signal matrix” \( M_{RP} \) and by \( \tilde{U} \) the corresponding orthonormal eigenvectors :

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

⇒ Intuition: \( \theta_1, \ldots, \theta_{K+1} \) largest \( K + 1 \) “true” eigenvalues of \( S_\gamma \).
**Theorem 1: Risk-Premium PCA under weak factor model**

Assumption 1 holds. The first $K$ largest eigenvalues $\hat{\theta}_i$, $i = 1, \ldots, K$ of $S_\gamma$ satisfy

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

The correlation of the estimated with the true factors converges to

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

with

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

\[ \hat{\Sigma}_{\hat{F}} = D_K^{1/2} \left( \begin{pmatrix} \rho_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_K \end{pmatrix} \right)^T \tilde{U}^T \begin{pmatrix} I_K & 0 \\ 0 & 0 \end{pmatrix} \tilde{U} \left( \begin{pmatrix} \rho_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_K \end{pmatrix} \right) + \left( \begin{pmatrix} 1 - \rho_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 - \rho_K^2 \end{pmatrix} \right) D_K^{1/2} \]

\[ D_K = \text{diag} \left( (\hat{\theta}_1, \ldots, \hat{\theta}_K) \right) \]
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.

$$\theta_i > \sigma^2_{F_i} + c\sigma^2_e$$

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

$$\frac{\partial \theta_i}{\partial \theta_i} > 0 \quad \frac{\partial \rho_i}{\partial \theta_i} > 0 \quad i = 1, ..., K$$

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.
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:

\[
\theta_{1,2} = \frac{1}{2} \sigma_F^2 + c\sigma_e^2 + (\mu^2 + c\sigma_e^2)(1 + \gamma)
\]

\[
\pm \frac{1}{2} \sqrt{(\sigma_F^2 + c\sigma_e^2 + (\mu^2 + c\sigma_e^2)(1 + \gamma))^2 - 4(1 + \gamma)c\sigma_e^2(\sigma_F^2 + \mu^2 + c\sigma_e^2)}
\]

The eigenvector of first eigenvalue \( \theta_1 \) 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:

\[
\hat{\text{Corr}}(F, \hat{F}) \xrightarrow{p} \frac{\rho_1}{\sqrt{\rho_1^2 + (1 - \rho_1^2) \left(\frac{(\theta_1 - (\sigma_F^2 + c\sigma_e^2))^2 + 1}{\mu^2\sigma_F^2(1+\gamma)}\right)}}
\]
Strong Factor Model

- Strong factors affect most assets: e.g. market factor
- $\frac{1}{N} \Lambda^T \Lambda$ 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 $\Lambda$ on $X$.

- **RP-PCA under slightly stronger assumptions as in Bai (2003):**
  - Asymptotically $\hat{\Lambda}$ behaves like OLS regression of $FW$ on $XW$ with $W^2 = \left( I_T + \gamma \frac{11^T}{T} \right)$.
  - Asymptotically $\hat{F}$ behaves like OLS regression of $\Lambda$ on $X$.

### Asymptotic Expansion

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

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

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

with known rotation matrix $H$. 
Assumption 2: Strong Factor Model

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

\[
\begin{pmatrix}
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{t,i} \\
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_{t,i}
\end{pmatrix} \sim N(0, \Omega) \quad \Omega = \begin{pmatrix}
\Omega_{1,1} & \Omega_{1,2} \\
\Omega_{2,1} & \Omega_{2,2}
\end{pmatrix}
\]
Theorem 2: Strong Factor Model

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

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

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

$$\cdot \left( \Sigma_F + (\gamma + 1)\mu_F\mu_F^T \right)^{-1}$$

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 $\gamma$.
- The asymptotic distribution of the common component depends on $\gamma$ if and only if $\frac{N}{T}$ does not go to zero. For $\frac{T}{N} \to 0$

$$\sqrt{T} \left( \hat{C}_{t,i} - C_{t,i} \right) \xrightarrow{D} N \left( 0, F_t^T \Phi F_t \right)$$