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

Martin Lettau $^1$  Markus Pelger $^2$

$^1$UC Berkeley

$^2$Stanford University

November 30th 2018

NBER Asset Pricing Meeting
Motivation

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

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

\[
X_{nt} = F_t \mathbf{B}_n^T + e_{nt} \quad n = 1, \ldots, N \quad t = 1, \ldots, T
\]

\( \iff \)

\[
\begin{pmatrix}
X \end{pmatrix}_{T \times N} = \begin{pmatrix}
F \end{pmatrix}_{T \times K} \begin{pmatrix} 
\mathbf{B}^T \end{pmatrix}_{K \times N} + \begin{pmatrix} 
e \end{pmatrix}_{T \times N}
\]

- \( T \): time-series observation (large)
- \( N \): test assets (large)
- \( K \): systematic factors (fixed)
- \( F, B \) and \( e \) are unknown
A statistical model of asset returns: Systematic factors

- Systematic and non-systematic risk:
  \[ \text{Var}(X) = \mathbf{B} \text{Var}(\mathbf{F}) \mathbf{B}^T + \text{Var}(\mathbf{e}) \]
  - Systematic factors explain large portion of variance
  - Idiosyncratic risk only weakly correlated
- Estimation via standard PCA: Minimize the unexplained variance
  \[ \min_{\mathbf{B}, \mathbf{F}} \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} (X_{nt} - \mathbf{F}_t \mathbf{B}_n^T)^2 \]
  - Easy to estimate: Eigenvectors/values of covariance matrix of \( X \)
    - PCA factors capture common co-movements
    - ... but usually do not fit cross-sectional mean returns and low SR
    - Example: Industry factors
A statistical model of asset returns: Risk premia

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

  \[ E[X_n] = B_n E[F] \]

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

- **Estimation**: Minimize cross-sectional pricing error

  \[
  \min_{B,F} \frac{1}{N} \sum_{n=1}^{N} \left( \bar{X}_n - \bar{F} B_i^T \right)^2
  \]

  with \( \bar{X}_n = \frac{1}{T} \sum_{t=1}^{T} X_{nt} \)
**Risk-Premium PCA (RP-PCA) estimator**

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

\[
\min_{B,F} \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} (X_{nt} - F_t B_n^T)^2 + \gamma \frac{1}{N} \sum_{n=1}^{N} (\bar{X}_n - \bar{F} B_n^T)^2
\]

- \( \gamma \) is the weight on the APT mean restriction
- Estimation: Apply standard PCA to the matrix with overweighted mean
  \[
  \frac{1}{T} x^T x + \gamma \bar{x}\bar{x}^T.
  \]
- Special case: \( \gamma = -1 \): (Standard) PCA on covariance matrix
Interpretation of RP-PCA

1. Combines variation and pricing error criterion functions:
   - Protects against spurious factor with vanishing loadings as it requires the time-series errors to be small as well.

2. Penalized PCA: Search for factors explaining the time-series but penalizes low Sharpe-ratios (consistent with APT)

3. Information interpretation: (GMM interpretation)
   - PCA of a covariance matrix uses only the second moment.
   - RP-PCA combines first and second moments efficiently.

4. Signal-strengthening: Intuitively the matrix $\frac{1}{T}X^T X + \gamma \bar{X} \bar{X}^T$ converges to $B \left( \text{Var}(F) + (1 + \gamma)E[F]E[F]^T \right) B^T + \text{Var}(e)$

   The signal of weak factors with a small variance can be “pushed up” by their mean with the right $\gamma$. 
Theoretical results

- Types of factors:
  - **Strong** factors affect all assets: MKT
  - **Weak** factors affect either some assets “strongly or all asset weakly”: Long-short factors

- Theory: Companion paper (Lettau and Pelger (2018))
  - **Strong** factors: RP-PCA more efficient
    - RP-PCA with $\gamma = 0$ is optimal but PCA works as well
  - **Weak** factors: RP-PCA dominates PCA
    - RP-PCA is able to detect factors that are missed by PCA
    - Optimal $\gamma \approx 10 – 20$
    - Novel results based on random matrix theory for non-zero means

- Paper:
  - Criterion to select number of factors
  - Extensive simulation study, in-sample and out-of-sample
### Illustration (Size and accrual)

<table>
<thead>
<tr>
<th>Model</th>
<th>SR</th>
<th>RMS $\alpha$</th>
<th>Idio. Var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP-PCA</td>
<td>0.21</td>
<td>0.11</td>
<td>6.75</td>
</tr>
<tr>
<td>PCA</td>
<td>0.11</td>
<td>0.14</td>
<td>6.72</td>
</tr>
<tr>
<td>FF-long/sort</td>
<td>0.11</td>
<td>0.12</td>
<td>7.11</td>
</tr>
</tbody>
</table>

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

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

RP-PCA:
SR₁ = 0.11
SR₂ = 0.03
SR₃ = 0.23

PCA:
SR₁ = 0.10
SR₂ = 0.02
SR₃ = 0.11
Large Cross-section of Single-Sorted Portfolios

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

### Single-sorted portfolios

<table>
<thead>
<tr>
<th>Model</th>
<th>Out-of-sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
</tr>
<tr>
<td>RP-PCA</td>
<td>0.45</td>
</tr>
<tr>
<td>PCA</td>
<td>0.17</td>
</tr>
<tr>
<td>Fama-French 5 factors</td>
<td>0.31</td>
</tr>
</tbody>
</table>

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

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

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

- **RP-PCA:** 5th factor (green) has big effect on SR
- **PCA:** higher order factors contribute less to SR
- **Extreme deciles capture most of the pricing information**
Choice of $\gamma$: Maximal Sharpe-ratio

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

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

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

- RP-PCA captures (almost) as much common variation as PCA factors
RMS of Pricing Errors $\alpha$’s: $N = 370$

- Characteristics sorted according to Sharpe-ratios of LS portfolios
- High SR characteristics significantly better priced by RP-PCA
Composition of factors

- Loading weights within deciles for all characteristics.
  - Almost all weights on extreme deciles.
Stochastic Discount Factors for $N = 74$

Order portfolios by SR! (top RP-PCA, bottom PCA)
**RP-PCA vs. PCA Factors**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Mean</th>
<th>Variance</th>
<th>SR</th>
<th>Mean</th>
<th>Variance</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>11.67</td>
<td>7295.35</td>
<td>0.14</td>
<td>11.56</td>
<td>7387.22</td>
<td>0.13</td>
</tr>
<tr>
<td>2.</td>
<td>2.65</td>
<td>222.62</td>
<td>0.18</td>
<td>1.66</td>
<td>241.03</td>
<td>0.11</td>
</tr>
<tr>
<td>3.</td>
<td>0.46</td>
<td>213.34</td>
<td>0.03</td>
<td>0.23</td>
<td>207.49</td>
<td>0.02</td>
</tr>
<tr>
<td>4.</td>
<td>2.40</td>
<td>125.92</td>
<td>0.21</td>
<td>1.52</td>
<td>132.57</td>
<td>0.13</td>
</tr>
<tr>
<td>5.</td>
<td>2.76</td>
<td>39.10</td>
<td>0.44</td>
<td>0.78</td>
<td>49.30</td>
<td>0.11</td>
</tr>
</tbody>
</table>

- Normalize loadings $B^T B = I_K$.
- Fifth RP-PCA factor weak and high SR.
**Interpretation of factors**

<table>
<thead>
<tr>
<th>RP-PCA</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Factor long-only market proxy</td>
<td>long-only market proxy</td>
</tr>
<tr>
<td>2. Factor value/growth + interactions</td>
<td>value/growth</td>
</tr>
<tr>
<td>3. Factor momentum+profitability</td>
<td>?</td>
</tr>
<tr>
<td>4. Factor momentum+momentum interaction</td>
<td>momentum</td>
</tr>
<tr>
<td>5. Factor high SR long/short portfolios</td>
<td>?</td>
</tr>
</tbody>
</table>

Note: Factors are comprised mostly of “classic” anomaly portfolios
Generalized Correlation with Ad-Hoc Long/Short Portfolios

![Graphs showing generalized correlations with increasing number of long-short factors.](image)

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

- First LS-factor is the market factor and LS-factors added incrementally based on the largest accumulative absolute loading.
  
  ⇒ Ad-hoc long-short factors do not span statistical factors.
Time-stability of loadings

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

⇒ RP-PCA stable over time
Next steps

- So far: Pre-sorted portfolios, constant factor-weights and loadings
- Goal: Construct factors using individual stocks without any pre-sorts
- Obstacle: Factor loadings of individual stocks are likely to vary over time
- Example: Momentum exposure
- Current work: Time-varying loadings

General idea behind all time-varying factor models

1. Create portfolios = projection based on characteristics
2. Estimate constant loading model on projected data
3. Invert model to obtain loadings that vary with characteristics

Our approach

- More general portfolios (projection) based on decision trees
- RP-PCA to extract factors on tree portfolios
- Invert model with second dimension reduction
Conclusion

Methodology

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

Empirical Results

- RP-PCA dominates standard PCA
- RP-PCA higher-order factors with high SRs
- RP-PCA loadings put more weight on high-SR characteristics
- Robust in out-of-sample estimation
- 5 economically meaningful asset pricing factors
- Redundancy in characteristic information
Factors 1: Long-only ("Mkt")

- Factor 1: Long in (almost) all portfolios
Factor 2: Value and value-interaction

- **RP-PCA**: Long/short in value and value-interaction portfolios
- **PCA**: Mostly value portfolios
Factor 3: Momentum

- RP-PCA: Momentum-related portfolios
- PCA: No clear pattern
Factor 4: Momentum-Interaction

- RP-PCA and PCA: Momentum and momentum-interaction portfolios
Factor 5: High SR

Note: Order portfolios by SR instead of categories!

RP-PCA: Long in highest SR portfolios
PCA: Asset Turnover and Profitability
Individual stocks

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

- Stock price data from 01/1972 to 12/2016 ($N = 270$ and $T = 500$)
  - 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
Time-stability of loadings of individual stocks

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

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

- RP-PCA has smaller pricing errors.
Single-sorted portfolios

<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</td>
<td>α</td>
<td>Idio. Var.</td>
</tr>
<tr>
<td>RP-PCA</td>
<td>0.53</td>
<td>0.14</td>
<td>10.76%</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>0.24</td>
<td>0.14</td>
<td>10.66%</td>
<td></td>
</tr>
<tr>
<td>Fama-French 5</td>
<td>0.32</td>
<td>0.23</td>
<td>13.56%</td>
<td></td>
</tr>
</tbody>
</table>

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

- RP-PCA strongly dominates PCA and Fama-French 5 factors
- Results hold out-of-sample.
Portfolio data: Average Pricing Errors

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

- RP-PCA has smaller pricing errors out-of-sample.
RMS of TS $\alpha$’s: $N = 74$
All 370 portfolios: RP-PCA
All 370 portfolios: PCA
Number of factors

Onatski (2010): Eigenvalue-ratio test

Onatski (2010): Eigenvalue-ratio test

- **N=74**: Eigenvalue Differences
- **N=370**: Eigenvalue Differences

- **RP-PCA**: 5 factors
- **PCA**: 4 factors
Figure: Generalized correlations between loadings estimated on the whole time horizon $T = 650$ and a rolling window with 240.
Signal of factors: Existence of weak factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>PCA</th>
<th>RP-PCA ($\gamma = 10$)</th>
<th>FF5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_1$</td>
<td>8.05</td>
<td>8.05</td>
<td>8.00</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>0.27</td>
<td>0.27</td>
<td>0.21</td>
</tr>
<tr>
<td>$\sigma^2_3$</td>
<td>0.21</td>
<td>0.21</td>
<td>0.17</td>
</tr>
<tr>
<td>$\sigma^2_4$</td>
<td>0.14</td>
<td>0.14</td>
<td>0.03</td>
</tr>
<tr>
<td>$\sigma^2_5$</td>
<td>0.05</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td>$\sigma^2_6$</td>
<td>0.03</td>
<td>0.03</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table: Variance signal for different factors

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

$\Rightarrow$ Higher factors are weak.
Signal of factors: Existence of weak factors

![Eigenvalues](image)

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

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

⇒ Higher factors have weak variance but high mean signal.
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>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>3. Gen. Corr.</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>4. Gen. Corr.</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>5. Gen. Corr.</td>
<td>0.77</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.
Single-sorted portfolios: Maximal Sharpe-ratio

**Figure:** Maximal Sharpe-ratios.

⇒ Spike in Sharpe-ratio for 5 factors
Single-sorted portfolios: Pricing error

Figure: Root-mean-squared pricing errors.

⇒ RP-PCA has smaller out-of-sample pricing errors
Single-sorted portfolios: Idiosyncratic Variation

Figure: Unexplained idiosyncratic variation.

⇒ Unexplained variation similar for RP-PCA and PCA
## Interpreting factors: 5th proxy factor

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

<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 (12m) - mom12</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 - momrev</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 - noa</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 - 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 - prof</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) - roa</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) - roea</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 - season</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 - sgrowth</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 - shvol</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 - 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 sp</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 - strev</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 - valmom</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. - valmomprof</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 - valprof</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) - value</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) - valuem</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>

**Table:** Long-Short Portfolios of extreme deciles of 37 single-sorted portfolios from 07/1963 to 12/2017: Mean, standard deviation and Sharpe-ratio.
<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Mean</th>
<th>SR</th>
<th>Portfolio</th>
<th>Mean</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind. Rel. Rev. (L.V.)</td>
<td>1.33</td>
<td>0.44</td>
<td>Return on Assets (A)</td>
<td>0.21</td>
<td>0.05</td>
</tr>
<tr>
<td>Industry Mom. Rev.</td>
<td>1.18</td>
<td>0.33</td>
<td>Leverage</td>
<td>0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>Industry Rel. Reversals</td>
<td>1.00</td>
<td>0.24</td>
<td>Dividend/Price</td>
<td>0.20</td>
<td>0.03</td>
</tr>
<tr>
<td>Seasonality</td>
<td>0.81</td>
<td>0.20</td>
<td>Net Operating Assets</td>
<td>0.15</td>
<td>0.02</td>
</tr>
<tr>
<td>Value-Profitability</td>
<td>0.75</td>
<td>0.19</td>
<td>Investment/Capital</td>
<td>0.12</td>
<td>0.02</td>
</tr>
<tr>
<td>Momentum (12m)</td>
<td>1.28</td>
<td>0.18</td>
<td>Return on Book Equity (A)</td>
<td>0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>Value-Mom-Prof.</td>
<td>0.84</td>
<td>0.17</td>
<td>Gross Margins</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Investment/Assets</td>
<td>0.48</td>
<td>0.15</td>
<td>Share Volume</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Composite Issuance</td>
<td>0.45</td>
<td>0.13</td>
<td>Price</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Investment Growth</td>
<td>0.37</td>
<td>0.13</td>
<td>Sales Growth</td>
<td>0.04</td>
<td>0.00</td>
</tr>
</tbody>
</table>
## Extreme Deciles

<table>
<thead>
<tr>
<th></th>
<th>In-sample</th>
<th></th>
<th></th>
<th></th>
<th>Out-of-sample</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
<td>RMS α</td>
<td>Idio. Var.</td>
<td></td>
<td>SR</td>
<td>RMS α</td>
<td>Idio. Var.</td>
<td></td>
</tr>
<tr>
<td><strong>RP-PCA</strong></td>
<td>0.57</td>
<td>0.17</td>
<td>10.40%</td>
<td></td>
<td>0.50</td>
<td>0.15</td>
<td>12.06%</td>
<td></td>
</tr>
<tr>
<td><strong>PCA</strong></td>
<td>0.30</td>
<td>0.22</td>
<td>10.30%</td>
<td></td>
<td>0.24</td>
<td>0.20</td>
<td>11.98%</td>
<td></td>
</tr>
<tr>
<td><strong>RP-PCA Proxy</strong></td>
<td>0.58</td>
<td>0.17</td>
<td>10.40%</td>
<td></td>
<td>0.50</td>
<td>0.15</td>
<td>11.97%</td>
<td></td>
</tr>
<tr>
<td><strong>PCA Proxy</strong></td>
<td>0.33</td>
<td>0.22</td>
<td>11.09%</td>
<td></td>
<td>0.27</td>
<td>0.18</td>
<td>12.10%</td>
<td></td>
</tr>
<tr>
<td><strong>Fama-French 5</strong></td>
<td>0.32</td>
<td>0.30</td>
<td>13.56%</td>
<td></td>
<td>0.31</td>
<td>0.26</td>
<td>13.66%</td>
<td></td>
</tr>
</tbody>
</table>

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

- RP-PCA strongly dominates PCA and Fama-French 5 factors
- Results hold out-of-sample.
Interpreting factors: Generalized correlations with proxies

<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>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>3. Gen. Corr.</td>
<td>0.95</td>
<td>0.97</td>
</tr>
<tr>
<td>4. Gen. Corr.</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td>5. Gen. Corr.</td>
<td>0.71</td>
<td>0.86</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.
Extreme Deciles: Maximal Sharpe-ratio

**Figure:** Maximal Sharpe-ratios.

⇒ Spike in Sharpe-ratio for 5 factors
Extreme Deciles: Pricing error

**Figure:** Root-mean-squared pricing errors.

⇒ RP-PCA has smaller out-of-sample pricing errors
Figure: Unexplained idiosyncratic variation.

⇒ Unexplained variation similar for RP-PCA and PCA
Extreme Deciles: Maximal Sharpe-ratio

Figure: Maximal Sharpe-ratios for different RP-weights $\gamma$ and number of factors $K$.
## Interpreting factors: 5th proxy factor

<table>
<thead>
<tr>
<th>5. Proxy RP-PCA</th>
<th>Weights</th>
<th>5. Proxy PCA</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value 10</td>
<td>1.93</td>
<td>Value-Profitability 10</td>
<td>1.25</td>
</tr>
<tr>
<td>Industry Rel. Reversal 10</td>
<td>1.39</td>
<td>Asset Turnover 10</td>
<td>1.15</td>
</tr>
<tr>
<td>Price 1</td>
<td>1.31</td>
<td>Profitability 10</td>
<td>0.95</td>
</tr>
<tr>
<td>Industry Rel. Reversal (LV) 10</td>
<td>1.26</td>
<td>Sales/Price 10</td>
<td>0.95</td>
</tr>
<tr>
<td>Long Run Reversals 10</td>
<td>1.25</td>
<td>Long Run Reversals 10</td>
<td>0.86</td>
</tr>
<tr>
<td>Short Run Reversals 1</td>
<td>-1.22</td>
<td>Value-Profitability 1</td>
<td>-0.98</td>
</tr>
<tr>
<td>Industry Rel. Reversal (LV) 1</td>
<td>-1.34</td>
<td>Profitability 1</td>
<td>-1.51</td>
</tr>
<tr>
<td>Industry Rel. Reversal 1</td>
<td>-1.37</td>
<td>Asset Turnover 1</td>
<td>-1.89</td>
</tr>
</tbody>
</table>
Interpreting factors: Composition of proxies

<table>
<thead>
<tr>
<th>RP-PCA</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>divp 10</td>
<td>1.53</td>
</tr>
<tr>
<td>mom12 1</td>
<td>2.04</td>
</tr>
<tr>
<td>size 10</td>
<td>2.14</td>
</tr>
<tr>
<td>valuem 10</td>
<td>1.93</td>
</tr>
<tr>
<td>mom 10</td>
<td>1.99</td>
</tr>
<tr>
<td>mom12 10</td>
<td>1.90</td>
</tr>
<tr>
<td>price 1</td>
<td>2.52</td>
</tr>
<tr>
<td>ivol 1</td>
<td>-1.46</td>
</tr>
<tr>
<td>mom 1</td>
<td>1.53</td>
</tr>
<tr>
<td>mom12 1</td>
<td>-1.51</td>
</tr>
<tr>
<td>indmomrev 10</td>
<td>1.90</td>
</tr>
<tr>
<td>valuem 10</td>
<td>-2.32</td>
</tr>
<tr>
<td>mom 10</td>
<td>2.39</td>
</tr>
<tr>
<td>mom12 10</td>
<td>1.84</td>
</tr>
<tr>
<td>indmomrev 10</td>
<td>1.26</td>
</tr>
<tr>
<td>valuem 10</td>
<td>-2.32</td>
</tr>
<tr>
<td>mom 10</td>
<td>2.39</td>
</tr>
<tr>
<td>mom12 10</td>
<td>2.12</td>
</tr>
<tr>
<td>irrev 10</td>
<td>1.06</td>
</tr>
<tr>
<td>mom 10</td>
<td>2.39</td>
</tr>
<tr>
<td>mom12 10</td>
<td>2.12</td>
</tr>
<tr>
<td>mom12 1</td>
<td>1.88</td>
</tr>
<tr>
<td>mom 1</td>
<td>1.53</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.
Figure: Time-varying rotated loadings for the first six factors. Loadings are estimated on a rolling window with 240 months.
Extreme Deciles: Time-stability: Generalized correlations

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

- Data
  - Monthly return data from 07/1963 to 12/2017 (\( T = 650 \))
  - 13 double sorted portfolios (consisting of 25 portfolios) from Kenneth French’s website

- Factors
  1. PCA: \( K = 3 \)
  2. RP-PCA: \( K = 3 \) and \( \gamma = 10 \)
  3. FF-Long/Short factors: market + two specific anomaly long-short factors
Sharpe-ratios and pricing errors (in-sample)

<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>0.22</td>
<td>0.21</td>
<td>0.13</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td><strong>BM and Investment</strong></td>
<td>0.18</td>
<td>0.17</td>
<td>0.24</td>
<td>0.11</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>BM and Profits</td>
<td>0.21</td>
<td>0.20</td>
<td>0.24</td>
<td>0.11</td>
<td>0.12</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>Size and Accrual</strong></td>
<td>0.24</td>
<td>0.13</td>
<td>0.21</td>
<td>0.12</td>
<td>0.14</td>
<td>0.12</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>
</tr>
<tr>
<td>Size and Investment</td>
<td>0.29</td>
<td>0.26</td>
<td>0.21</td>
<td>0.11</td>
<td>0.11</td>
<td>0.22</td>
</tr>
<tr>
<td>Size and Profits</td>
<td>0.21</td>
<td>0.21</td>
<td>0.25</td>
<td>0.06</td>
<td>0.06</td>
<td>0.16</td>
</tr>
<tr>
<td>Size and Momentum</td>
<td>0.21</td>
<td>0.19</td>
<td>0.18</td>
<td>0.15</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td>Size and ST-Reversal</td>
<td>0.27</td>
<td>0.25</td>
<td>0.24</td>
<td>0.16</td>
<td>0.17</td>
<td>0.35</td>
</tr>
<tr>
<td>Size and Idio. Vol.</td>
<td>0.33</td>
<td>0.31</td>
<td>0.32</td>
<td>0.15</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>Size and Total Vol.</td>
<td>0.32</td>
<td>0.30</td>
<td>0.31</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>Profits and Invest.</td>
<td>0.26</td>
<td>0.24</td>
<td>0.30</td>
<td>0.11</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>Size and LT-Reversal</td>
<td>0.19</td>
<td>0.18</td>
<td>0.16</td>
<td>0.12</td>
<td>0.13</td>
<td>0.16</td>
</tr>
</tbody>
</table>
## Sharpe-ratios and pricing errors (out-of-sample)

<table>
<thead>
<tr>
<th></th>
<th>Sharpe-Ratio</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>
</tr>
<tr>
<td>Size and BM</td>
<td>0.22</td>
<td>0.18</td>
<td>0.16</td>
<td>0.18</td>
<td>0.19</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>BM and Investment</td>
<td><strong>0.18</strong></td>
<td><strong>0.15</strong></td>
<td><strong>0.24</strong></td>
<td><strong>0.16</strong></td>
<td><strong>0.17</strong></td>
<td><strong>0.17</strong></td>
<td></td>
</tr>
<tr>
<td>BM and Profits</td>
<td>0.19</td>
<td>0.17</td>
<td>0.23</td>
<td>0.17</td>
<td>0.17</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Size and Accrual</td>
<td><strong>0.24</strong></td>
<td><strong>0.09</strong></td>
<td><strong>0.11</strong></td>
<td><strong>0.11</strong></td>
<td><strong>0.14</strong></td>
<td><strong>0.12</strong></td>
<td></td>
</tr>
<tr>
<td>Size and Beta</td>
<td>0.21</td>
<td>0.20</td>
<td>0.16</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>Size and Investment</td>
<td>0.29</td>
<td>0.23</td>
<td>0.17</td>
<td>0.13</td>
<td>0.14</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Size and Profits</td>
<td>0.21</td>
<td>0.20</td>
<td>0.20</td>
<td>0.10</td>
<td>0.10</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Size and Momentum</td>
<td>0.17</td>
<td>0.12</td>
<td>0.08</td>
<td>0.18</td>
<td>0.18</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Size and ST-Reversal</td>
<td>0.21</td>
<td>0.17</td>
<td>0.23</td>
<td>0.22</td>
<td>0.23</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Size and Idio. Vol.</td>
<td>0.36</td>
<td>0.30</td>
<td>0.28</td>
<td>0.17</td>
<td>0.18</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>Size and Total Vol.</td>
<td>0.34</td>
<td>0.28</td>
<td>0.27</td>
<td>0.19</td>
<td>0.20</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Profits and Invest.</td>
<td>0.31</td>
<td>0.25</td>
<td>0.29</td>
<td>0.13</td>
<td>0.15</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>Size and LT-Reversal</td>
<td>0.11</td>
<td>0.10</td>
<td>0.04</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td></td>
</tr>
</tbody>
</table>
Maximal Sharpe ratio (Size and accrual)

**Figure**: Maximal Sharpe-ratio by adding factors incrementally.

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

⇒ RP-PCA detects 3rd factor (accrual) for $\gamma > 10$. 
Effect of Risk-Premium Weight $\gamma$

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

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

- Large-dimensional factor models with strong factors
  - Bai (2003): Distribution theory
  - Bai and Ng (2017): Robust PCA
  - Fan et al. (2016): Projected PCA for time-varying loadings
  - Kelly et al. (2017): Instrumented PCA for time-varying loadings
  - Pelger (2016), Aït-Sahalia and Xiu (2015): High-frequency

- Large-dimensional factor models with weak factors
  (based on random matrix theory)
  - Onatski (2012): Phase transition phenomena
  - Benauch-Georges and Nadakuditi (2011): Perturbation of large random matrices

- Asset-pricing factors
  - Feng, Giglio and Xiu (2017): Factor selection with double-selection LASSO
  - Kozak, Nagel and Santosh (2017): Bayesian shrinkage
The Model

Approximate Factor Model

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

$$X_{t,i} = F_t^\top \Lambda_i + e_{t,i} \quad i = 1, \ldots, N \quad t = 1, \ldots, T$$

- Matrix notation

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

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

- $F$, $\Lambda$ and $e$ are unknown
Approximate Factor Model

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

$$\text{Var}(X) = \Lambda \text{Var}(F)\Lambda^T + \text{Var}(e)$$

⇒ 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:

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

⇒ 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:

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

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

Minimize jointly the unexplained variance and pricing error

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

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: Objective function

Variation objective function:

Minimize the unexplained variation:

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

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

- 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

⇒ Motivation for PCA
The Model: Objective function

Pricing objective function:

Minimize cross-sectional expected pricing error:

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

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

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

- The objective function is minimized by the eigenvectors of the largest eigenvalues of \( \frac{1}{NT} X^\top (I_T + \frac{\gamma}{T} \mathbb{1} \mathbb{1}^\top) X \).

- \( \hat{\Lambda} \) estimator for loadings: proportional to eigenvectors of the first \( K \) eigenvalues of \( \frac{1}{NT} X^\top (I_T + \frac{\gamma}{T} \mathbb{1} \mathbb{1}^\top) X \)

- \( \hat{F} \) 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:

\[
\begin{align*}
\min_{\Lambda,F} \ & \frac{1}{NT} \text{trace}(Q^T (X - F\Lambda^T)^T (X - F\Lambda^T)Q) \\
& + \gamma \frac{1}{N} \text{trace} (\mathbb{1}^T (X - F\Lambda^T)QQ^T (X - F\Lambda^T)^T \mathbb{1}) \\
= \min_{\Lambda} \text{trace} \left( M_\Lambda Q^T X^T \left( I + \frac{\gamma}{T} \mathbb{1}\mathbb{1}^T \right) XQM_\Lambda \right) \quad \text{s.t. } F = X(\Lambda^T \Lambda)^{-1} \Lambda^T
\end{align*}
\]

- Cross-sectional weighting matrix \(Q\)
- Factors and loadings can be estimated by applying PCA to \(Q^T X^T \left( I + \frac{\gamma}{T} \mathbb{1}\mathbb{1}^T \right) XQ\).
- 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
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
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 mean zero
  ⇒ RP-PCA implicitly uses non-zero means of random variables
  ⇒ New tools necessary!
Assumption 1: Weak Factor Model

1. **Rate:** Assume that \( \frac{N}{T} \to c \) 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

Definition: Weak Factor Model

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

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

- $B$-function

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

Intuition: Weak Factor Model

- “Signal” matrix for PCA of covariance matrix ($\gamma = -1$):
  \[
  \Sigma_F + c\sigma_e^2 I_K
  \]
  $K$ largest eigenvalues $\theta_{1,\text{PCA}}, ..., \theta_{K,\text{PCA}}$ measure strength of signal

- “Signal” matrix for RP-PCA:
  \[
  \begin{pmatrix}
  \Sigma_F + c\sigma_e^2 & \Sigma_F^{1/2} \mu_F (1 + \tilde{\gamma}) \\
  \mu_F^\top \Sigma_F^{-1/2} (1 + \tilde{\gamma}) & (1 + \gamma)(\mu_F^\top \mu_F + c\sigma_e^2)
  \end{pmatrix}
  \]
  $(1 + \tilde{\gamma})^2 = 1 + \gamma$
  $K$ largest eigenvalues $\theta_{1,\text{RP-PCA}}, ..., \theta_{K,\text{RP-PCA}}$ measure strength of signal

- RP-PCA signal matrix is “close” to
  \[
  \Sigma_F + (1 + \gamma)\mu_F\mu_F^\top + c\sigma_e^2 I_K
  \]
Weak Factor Model

**Theorem 1: Risk-Premium PCA under weak factor model**

Assumption 1 holds. The first $K$ largest eigenvalues $\hat{\theta}_i$, $i = 1, ..., K$ of $\frac{1}{T}X^T \left( I_T + \gamma \frac{11^T}{T} \right) X$ satisfy

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

The correlation of the estimated with the true factors converges to

$$\text{Corr}(F, \hat{F}) \xrightarrow{p} \begin{pmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \rho_K \end{pmatrix}$$

with

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

Optimal choice or risk premium weight $\gamma$

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

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

$$\theta_{i}^{\text{RP-PCA}} > \theta_{i}^{\text{PCA}}$$

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

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

- The rotation matrices satisfy $\bar{U}^\top \bar{U} \leq I_K$ and $\bar{V}^\top \bar{V} \leq I_K$.

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

$\Rightarrow$ Based on closed-form expression choose optimal RP-weight $\gamma$.
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): (up to rotation)
  - Asymptotically $\hat{\Lambda}$ behaves like OLS regression of $F$ on $X$.
  - Asymptotically $\hat{F}$ behaves like OLS regression of $\Lambda$ on $X^T$.
- RP-PCA under slightly stronger assumptions as in Bai (2003):
  - Asymptotically $\hat{\Lambda}$ behaves like OLS regression of $FW$ on $XW$ with
    $W^2 = \left( I_T + \gamma \frac{\mathbb{1} \mathbb{1}^T}{T} \right)$ and $\mathbb{1}$ 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:** $\Lambda^\top \Lambda / N \xrightarrow{p} I_K$ and all loadings are bounded.
4. **Residuals:** $e = \epsilon \Sigma$ with $\epsilon_{t,i} \sim N(0, 1)$. All elements and all row sums of $\Sigma$ are bounded.
Simplified Strong Factor Model

Proposition: Simplified Strong Factor Model

Assumption 2 holds. Then:

1. The factors and loadings can be estimated consistently.
2. The asymptotic distribution of the factors is not affected by $\gamma$.
3. The asymptotic distribution of the loadings is given by
   \[ \sqrt{T} \left( H\hat{\Lambda}_i - \Lambda_i \right) \xrightarrow{D} N(0, \Omega_i) \]
   \[ \Omega_i = \sigma^2_e \left( \Sigma_F + (1 + \gamma)\mu_F\mu_F^T \right)^{-1} \left( \Sigma_F + (1 + \gamma)^2\mu_F\mu_F^T \right) \left( \Sigma_F + (1 + \gamma)\mu_F\mu_F^T \right)^{-1} \]
   \[ E[e_{t,i}^2] = \sigma^2_e, \quad H \text{ full rank matrix} \]
4. $\gamma = 0$ is optimal choice for smallest asymptotic variance.
   $\gamma = -1$, i.e. the covariance matrix, is not efficient.
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 $\sqrt{T} \left( H^T \hat{\Lambda}_i - \Lambda_i \right) \overset{D}{\to} N(0, \Phi)$

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

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 $\frac{T}{N} \to 0$ $\sqrt{T} \left( \hat{C}_{t,i} - C_{t,i} \right) \overset{D}{\to} N(0, F_t^T \Phi F_t)$
Time-varying loadings

Model with time-varying loadings

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

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

- Loadings are function of $L$ covariates $Z_{i,t-1,l}$ with $l = 1, ..., L$
  - 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 $\phi_m(.)$:

\[
g_k(Z_{i,t-1}) = \sum_{m=1}^{M} b_{m,k} \phi_m(Z_{i,t-1}) \quad \Rightarrow \quad g(Z_{t-1}) = B^\top \Phi(Z_{t-1})
\]

- 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 $Z_{t-1}$. 
### Simulation parameters

- Parameters as in the empirical application
- \( N = 370 \) and \( T = 650 \).
- Factors:
  - \( K = 4 \) or \( K = 1 \)
  - Factors \( F_t \sim N(\mu_F, \Sigma_F) \)
  - \( \Sigma_F = \text{diag}(5, 0.3, 0.1, \sigma_F^2) \) with \( \sigma_F^2 \in \{0.03, 0.05, 0.1\} \)
  - \( SR_F = (0.12, 0.1, 0.3, sr) \) with \( sr \in \{0.8, 0.5, 0.3, 0.2\} \)
- Loadings: \( \Lambda_i \sim N(0, I_K) \)
- Residuals: \( e_t \sim \epsilon_t \Sigma \) with empirical correlation matrix and \( \sigma^2_e = 1 \).
Figure: Sample paths of the cumulative returns of the first four factors and the estimated factor processes. The fourth factor has a variance $\sigma_F^2 = 0.03$ and Sharpe-ratio $sr = 0.5$. 
Simulation: Multifactor Model

Figure: Correlation of estimated with true factor.
Simulation: Multifactor Model

Figure: Maximal Sharpe-ratio of factors.
Simulation: Weak factor model prediction

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

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

Correlation of estimated with true factors for different variances and Sharpe-ratios of the factor and for different RP-weights $\gamma$. 
Sharpe-ratio for different variances and Sharpe-ratios of the factor and for different RP-weights \( \gamma \). The residuals have the empirical residual correlation matrix.