Estimating Latent Asset-Pricing Factors

Markus Pelger ¹  Martin Lettau ²

¹Stanford University
²UC Berkeley

September 4th 2018

IEOR-DRO Seminar
Columbia University
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
Motivation

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
Motivation

Contribution of this paper

Contribution

- This Paper: Estimation approach for finding risk factors
- Key elements of estimator:
  1. Statistical factors instead of pre-specified (and potentially miss-specified) factors
  2. Uses information from large panel data sets: Many assets with many time observations
  3. Searches for factors explaining asset prices (explain differences in expected returns) not only co-movement in the data
  4. Allows time-variation in factor structure
Motivation

Contribution of this paper

Results

- Asymptotic distribution theory for weak and strong factors
  \[ \Rightarrow \] No “blackbox approach”

- Estimator discovers “weak” factors with high Sharpe-ratios
  \[ \Rightarrow \] high Sharpe-ratio factors important for asset pricing and investment

- Estimator strongly dominates conventional approach (Principal Component Analysis (PCA))
  \[ \Rightarrow \] PCA does not find all high Sharpe-ratio factors

- Empirical results:
  - New factors much smaller pricing errors in- and out-of sample than benchmark (PCA, 5 Fama-French factors, etc.)
  - 2 times higher Sharpe-ratio then benchmark factors (PCA)
Motivation

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:

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

- Factors loadings
- Idiosyncratic

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):

$$Var(X) = \Lambda Var(F)\Lambda^\top + Var(e)$$

⇒ 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:

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

⇒ Systematic factors should explain the cross-section of expected returns
The Model: Estimation of Latent Factors

Conventional approach: PCA (Principal component analysis)

- Apply PCA to the sample covariance matrix:
  \[
  \frac{1}{T} X^\top X - \bar{X}\bar{X}^\top
  \]

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

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

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

- Apply PCA to a covariance matrix with overweighted mean
  \[
  \frac{1}{T} X^\top X + \gamma \bar{X}\bar{X}^\top \quad \gamma = \text{risk-premium weight}
  \]

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

- \( \hat{F} \) estimator for factors: \( \hat{F} = \frac{1}{N} X \hat{\Lambda} = X (\hat{\Lambda}^\top \hat{\Lambda})^{-1} \hat{\Lambda}^\top \).
The Model: Objective Function

Conventional PCA: Objective Function

Minimize the unexplained variance:

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

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

Minimize jointly the unexplained variance and pricing error

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

with $\bar{X}_i = \frac{1}{T} \sum_{t=1}^{T} X_{t,i}$ and $\bar{F} = \frac{1}{T} \sum_{t=1}^{T} F_t$ and risk-premium weight $\gamma$
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: (GMM interpretation)
   - PCA of a covariance matrix uses only the second moment but ignores first moment
   - Using more information leads to more efficient estimates. RP-PCA combines first and second moments efficiently.
Signal-strengthening: Intuitively the matrix $\frac{1}{T} X^T X + \gamma \bar{X} \bar{X}^T$ 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 $\gamma$. 

Interpretation of Risk-Premium-PCA (RP-PCA): continued
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 \( \gamma = 10 \)
  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 12/2017 (\( T = 650 \)) for \( N = 25 \) portfolios
  - Out-of-sample: Rolling window of 20 years (\( T = 240 \))

- Stochastic Discount Factor (SDF): maximum Sharpe-ratio portfolio

\[ R_{opt} = F \cdot \sum_{F}^{-1} \mu_{F} \]
Table: Maximal Sharpe-ratios, root-mean-squared pricing errors and unexplained idiosyncratic variation. $K = 3$ factors and $\gamma = 10$.

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

⇒ 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: Maximal Sharpe-ratio by adding factors incrementally.]

**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 $\gamma > 10$.  

---

*Illustration*
Strong vs. weak factor models

- **Strong factor model** $(\frac{1}{N} \Lambda^T \Lambda$ bounded)
  - Interpretation: strong factors affect most assets (proportional to $N$), e.g. market factor
  - Strong factors lead to exploding eigenvalues
  - $\Rightarrow$ RP-PCA always more efficient than PCA
  - $\Rightarrow$ optimal $\gamma$ relatively small

- **Weak factor model** $(\Lambda^T \Lambda$ bounded)
  - Interpretation: weak factors affect a smaller fraction of assets
  - Weak factors lead to large but bounded eigenvalues
  - $\Rightarrow$ RP-PCA detects weak factors which cannot be detected by PCA
  - $\Rightarrow$ optimal $\gamma$ relatively large
Weak Factor Model

- Weak factors either have a small variance or affect a smaller fraction of assets:
  \[ \Lambda^\top \Lambda \text{ bounded (after normalizing factor variances)} \]
- 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
  $\Rightarrow$ 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**

$\Rightarrow$ RP-PCA implicitly uses non-zero means of random variables

$\Rightarrow$ New tools necessary!
Weak Factor Model

Assumption 1: Weak Factor Model

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

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

   \[
   \hat{\mu}_F := \frac{1}{T} \sum_{t=1}^{T} F_t \xrightarrow{p} \mu_F \quad \hat{\Sigma}_F := \frac{1}{T} F_t F_t^\top \xrightarrow{p} \Sigma_F = \begin{pmatrix} \sigma_{F1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{FK}^2 \end{pmatrix}
   \]

3. **Loadings:** The column vectors of the loadings $\Lambda$ are orthogonally invariant and independent of $e$ 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 $b$. 

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:

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

- $B$-function

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

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

Intuition: Weak Factor Model

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

\[
\Sigma_F + c\sigma_e^2 l_K
\]

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

- "Signal" matrix for RP-PCA:

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

$(1 + \tilde{\gamma})^2 = 1 + \gamma$

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

- RP-PCA signal matrix is "close" to

\[
\Sigma_F + (1 + \gamma)\mu_F\mu_F^\top + c\sigma_e^2 l_K
\]
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
\[
\frac{1}{T}X^T \left( I_T + \gamma \frac{11^T}{T} \right) X
\]
satisfy
\[
\hat{\theta}_i \xrightarrow{p} \begin{cases} 
G^{-1} \left( \frac{1}{\theta_i} \right) & \text{if } \theta_i > \theta_{\text{crit}} = \lim_{z \downarrow b} \frac{1}{G(z)} \\
 b & \text{otherwise}
\end{cases}
\]

The correlation of the estimated with the true factors converges to
\[
\text{Corr}(F, \hat{F}) \xrightarrow{p} \begin{pmatrix} 
\rho_1 & 0 & \cdots & 0 \\
0 & \rho_2 & \cdots & 0 \\
0 & 0 & \ddots & \vdots \\
0 & \cdots & 0 & \rho_K
\end{pmatrix}
\]
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 $\hat{\rho}_{i}^{2}$ is strictly increasing in $\theta_{i}$.
- The rotation matrices satisfy $\tilde{U}^{\top} \tilde{U} \leq I_{K}$ and $\tilde{V}^{\top} \tilde{V} \leq I_{K}$.

$\Rightarrow \widehat{\text{Corr}}(F, \hat{F})$ is not necessarily an increasing function in $\theta$.
- $\Rightarrow$ Based on closed-form expression choose optimal RP-weight $\gamma$. 

\[ \begin{align*}
\]
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)
### 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**: $\Lambda^\top \Lambda / N \xrightarrow{P} I_K$ 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.
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

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

   $$\Omega_i = \sigma^2_{e_i} \left( \Sigma_F + (1 + \gamma) \mu_F \mu_F^T \right)^{-1} \left( \Sigma_F + (1 + \gamma)^2 \mu_F \mu_F^T \right) \left( \Sigma_F + (1 + \gamma) \mu_F \mu_F^T \right)^{-1}$$

   $$E[e_{t,i}^2] = \sigma^2_{e_i}, \quad H \text{ full rank matrix}$$

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

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

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

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 $\phi_m(.)$:

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

- 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 $Z_{t-1}$. 
Empirical Results

Single-sorted portfolios

Portfolio Data

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

#### 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 $\alpha$</td>
<td>Idio. Var.</td>
<td>SR</td>
</tr>
<tr>
<td>RP-PCA</td>
<td>0.53</td>
<td>0.14</td>
<td>10.76%</td>
<td>0.45</td>
</tr>
<tr>
<td>PCA</td>
<td>0.24</td>
<td>0.14</td>
<td>10.66%</td>
<td>0.17</td>
</tr>
<tr>
<td>Fama-French 5</td>
<td>0.32</td>
<td>0.23</td>
<td>13.56%</td>
<td>0.31</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 5 factors
Empirical Results

Single-sorted portfolios: Pricing error

![RMS α (In-sample)](image1)

![RMS α (Out-of-sample)](image2)

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

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

Single-sorted portfolios: Idiosyncratic Variation

Figure: Unexplained idiosyncratic variation.

⇒ Unexplained variation similar for RP-PCA and PCA
Empirical Results

Choice of $\gamma$: Maximal Sharpe-ratio

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

$\Rightarrow$ Strong increase in Sharpe-ratios for $\gamma \geq 10$. 
Signal of factors: Existence of weak factors

<table>
<thead>
<tr>
<th></th>
<th>PCA</th>
<th>RP-PCA ($\gamma = 10$)</th>
<th>FF5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1^2$</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>$\sigma_3^2$</td>
<td>0.21</td>
<td>0.21</td>
<td>0.17</td>
</tr>
<tr>
<td>$\sigma_4^2$</td>
<td>0.14</td>
<td>0.14</td>
<td>0.03</td>
</tr>
<tr>
<td>$\sigma_5^2$</td>
<td>0.05</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td>$\sigma_6^2$</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 $\frac{1}{N} \Lambda \Sigma_F \Lambda^\top$ normalized by the average idiosyncratic variance $\sigma_e^2 = \frac{1}{N} \sum_{i=1}^{N} \sigma_{e,i}^2$

$\Rightarrow$ Higher factors are weak.
Empirical Results

**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 $\sigma^2_e = \frac{1}{N} \sum_{i=1}^{N} \sigma^2_{e,i}$.
- **RIGHT:** Eigenvalues are normalized by the corresponding PCA ($\gamma = -1$) eigenvalues.

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

Number of factors

Onatski (2010): Eigenvalue-ratio test

- RP-PCA: 5 factors
- PCA: 4 factors
Empirical Results

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.
### 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>
Single-sorted portfolios: Maximal Sharpe-ratio

![Graph](attachment:graph.png)

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

- Extreme deciles are lowest and highest decile portfolio for each anomaly \((N = 74)\).

⇒ Extreme deciles capture most of the pricing information
Stochastic Discount Factor

All 370 portfolios: PCA

- Loading weights within deciles for all characteristics.
- ⇒ Almost all weights on extreme deciles.
Figure: Portfolio composition of highest Sharpe ratio portfolio (Stochastic Discount Factor) based on 5 RP-PCA factors.
Optimal Portfolio with RP-PCA (largest positions)

Figure: Largest portfolios in highest Sharpe ratio portfolio (Stochastic Discount Factor) based on 5 RP-PCA factors.
**Figure**: Portfolio composition of highest Sharpe ratio portfolio (Stochastic Discount Factor) based on 5 PCA factors.
Figure: Largest portfolios in highest Sharpe ratio portfolio (Stochastic Discount Factor) based on 5 PCA factors.
Order portfolios by SR! (top RP-PCA, bottom PCA)
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**

**Time-stability: Generalized correlations**

![Graph showing generalized correlations for RP-PCA and PCA](image)

**Figure:** Generalized correlations between loadings estimated on the whole time horizon \( T = 650 \) and a rolling window with 240.
**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.
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.
RMS of TS $\alpha$’s: $N = 370$
RMS of TS $\alpha$’s: $N = 74$
Single-sorted portfolios: Interpreting factors

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.

⇒ Long-Short Factors do not span statistical factors.
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!

<table>
<thead>
<tr>
<th>Factor 5 (sorted by SR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP-PCA10</td>
</tr>
<tr>
<td>PCA10</td>
</tr>
<tr>
<td>PCA1</td>
</tr>
</tbody>
</table>

RP-PCA: Long in highest SR portfolios
PCA: Asset Turnover and Profitability
## Interpretation of factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>RP-PCA</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2</td>
<td>long</td>
<td>long</td>
</tr>
<tr>
<td>3</td>
<td>value &amp; value interactions</td>
<td>value</td>
</tr>
<tr>
<td>4</td>
<td>momentum</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>momentum-interaction</td>
<td>momentum-interaction</td>
</tr>
<tr>
<td>6</td>
<td>high SR</td>
<td>asset turnover and profitability</td>
</tr>
</tbody>
</table>

Note: Factors are comprised mostly of “classic” anomaly portfolios
All 370 portfolios: RP-PCA
The Model: Objective function

Variation objective function:

Minimize the unexplained variation:

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

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

- 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

⇒ Motivation for PCA
The Model: Objective function

Pricing objective function:

Minimize cross-sectional expected pricing error:

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

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

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

- $\mathbb{1}$ is vector $T \times 1$ of 1’s and thus $\frac{F^T \mathbb{1}}{T}$ estimates factor mean

- Why not estimate factors with cross-sectional objective function?
  - Factors not identified
  - Spurious factor detection (Bryzgalova (2016))
The 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 \( \hat{C} = \hat{F} \hat{\Lambda}^\top \).
Simulation

Simulation parameters

- Parameters as in the empirical application
- $N = 370$ and $T = 650$.
- Factors:
  - $K = 4$ or $K = 1$
  - Factors $F_t \sim N(\mu_F, \Sigma_F)$
  - $\Sigma_F = \text{diag}(5, 0.3, 0.1, \sigma_F^2)$ 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 $\sigma_e^2 = 1$. 
Figure: Sample paths of the cumulative returns of the first four factors and the estimated factor processes. The fourth factor has a variance $\sigma^2_F = 0.03$ and Sharpe-ratio $sr = 0.5$. 
Simulation: Multifactor Model

1. Factor Corr. (IS) for $\sigma_f^2=0.03$
2. Factor Corr. (IS) for $\sigma_f^2=0.03$
3. Factor Corr. (IS) for $\sigma_f^2=0.03$
4. Factor Corr. (IS) for $\sigma_f^2=0.03$

1. Factor Corr. (OOS) for $\sigma_f^2=0.03$
2. Factor Corr. (OOS) for $\sigma_f^2=0.03$
3. Factor Corr. (OOS) for $\sigma_f^2=0.03$
4. Factor Corr. (OOS) for $\sigma_f^2=0.03$

1. Factor Corr. (IS) for $\sigma_f^2=0.1$
2. Factor Corr. (IS) for $\sigma_f^2=0.1$
3. Factor Corr. (IS) for $\sigma_f^2=0.1$
4. Factor Corr. (IS) for $\sigma_f^2=0.1$

1. Factor Corr. (OOS) for $\sigma_f^2=0.1$
2. Factor Corr. (OOS) for $\sigma_f^2=0.1$
3. Factor Corr. (OOS) for $\sigma_f^2=0.1$
4. Factor Corr. (OOS) for $\sigma_f^2=0.1$

**Figure**: Correlation of estimated with true factor.
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 $\theta_i > \sigma_{\text{crit}}^2$ 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.
The Model: Objective function

Weighted Combined objective function:

Straightforward extension to weighted objective function:

\[
\min_{\Lambda,F} \frac{1}{NT} \text{trace}(Q^T (X - F\Lambda^T)^T (X - F\Lambda^T) Q) \\
+ \gamma \frac{1}{N} \text{trace} \left( (1^T (X - F\Lambda^T) QQ^T (X - F\Lambda^T)^T 1^T \right) \\
= \min_{\Lambda} \text{trace} \left( M_\Lambda Q^T X^T \left( I + \frac{\gamma}{T} 11^T \right) XQ M_\Lambda \right) \quad \text{s.t. } F = X(\Lambda^T \Lambda)^{-1} \Lambda^T
\]

- Cross-sectional weighting matrix \( Q \)
- Factors and loadings can be estimated by applying PCA to \( Q^T X^T \left( I + \frac{\gamma}{T} 11^T \right) XQ \).
- 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
Weak Factor Model

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

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

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

The correlation between the estimated and true factors converges to

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

with

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

Example: One-factor model

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

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

and has the eigenvalues:

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

The eigenvector of first eigenvalue $\theta_1$ has the components

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

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

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

\[
\hat{\text{Corr}}(F, \hat{F}) \overset{p}{\rightarrow} \frac{\rho_1}{\sqrt{\rho_1^2 + (1 - \rho_1^2)(\theta_1 - (\sigma_F^2 + c\sigma_e^2))^2 + 1}} \frac{\theta_1 - (\sigma_F^2 + c\sigma_e^2))^2 + 1}{\mu^2 \sigma_F^2 (1 + \gamma)}
\]
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):
  - 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. $\sqrt{N} \left( H^{T^{-1}} \hat{F}_t - F_t \right) = \left( \frac{1}{N} \Lambda^T \Lambda \right)^{-1} \frac{1}{\sqrt{N}} \Lambda^T e_t^T + O_p \left( \frac{\sqrt{N}}{T} \right) + o_p(1)$

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

\[
\left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{t,i} \right) \xrightarrow{D} N(0, \Omega) \quad \Omega = \begin{pmatrix} \Omega_{1,1} & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{pmatrix}
\]
**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 $\sqrt{T} \left( H^T \hat{\Lambda}_i - \Lambda_i \right) \overset{D}{\to} N(0, \Phi)$

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

- 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 $\frac{T}{N} \to 0$

$$
\sqrt{T} \left( \hat{C}_{t,i} - C_{t,i} \right) \overset{D}{\to} N \left( 0, F_t^T \Phi F_t \right)
$$
Extreme deciles of single-sorted portfolios

### Portfolio Data

- Monthly return data from 07/1963 to 12/2017 ($T = 650$) for $N = 74$ 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 ($N = 74$).

Factors:

1. **RP-PCA**: $K = 5$ and $\gamma = 10$.
2. **PCA**: $K = 5$
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 12/2017: Mean, standard deviation and Sharpe-ratio.

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

<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>0.57</td>
<td>0.17</td>
</tr>
<tr>
<td>PCA</td>
<td>0.30</td>
<td>0.22</td>
</tr>
<tr>
<td>RP-PCA Proxy</td>
<td>0.58</td>
<td>0.17</td>
</tr>
<tr>
<td>PCA Proxy</td>
<td>0.33</td>
<td>0.22</td>
</tr>
<tr>
<td>Fama-French 5</td>
<td>0.32</td>
<td>0.30</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.
Extreme Deciles: Number of factors

Onatski (2010): Eigenvalue-ratio test

![Eigenvalue Differences Graph]

- **RP-PCA**: 5 factors
- **PCA**: 4 factors
Extreme Deciles: Maximal Sharpe-ratio

Figure: Maximal Sharpe-ratios.

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

Figure: Root-mean-squared pricing errors.

⇒ RP-PCA has smaller out-of-sample pricing errors
Extreme Deciles: Idiosyncratic Variation

Figure: Unexplained idiosyncratic variation.

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

Figure: Maximal Sharpe-ratios for different RP-weights $\gamma$ 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>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.
## 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></th>
<th>RP-PCA</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>divp 10</td>
<td>1.53</td>
<td>valuem 10</td>
</tr>
<tr>
<td>growth 1</td>
<td>-1.46</td>
<td>divp 10</td>
</tr>
<tr>
<td>igrowth 1</td>
<td>-1.51</td>
<td>mom12 10</td>
</tr>
<tr>
<td>ep 1</td>
<td>-1.53</td>
<td>mom 10</td>
</tr>
<tr>
<td>invcap 1</td>
<td>-1.69</td>
<td>indmomrev 10</td>
</tr>
<tr>
<td>shvol 1</td>
<td>-1.72</td>
<td>vol 10</td>
</tr>
<tr>
<td>mom12 1</td>
<td>-2.32</td>
<td>price 1</td>
</tr>
<tr>
<td>ivol 1</td>
<td>-2.48</td>
<td>mom12 1</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 of loadings

**Figure:** Time-varying rotated loadings for the first six factors. Loadings are estimated on a rolling window with 240 months.
Figure: Generalized correlations between loadings estimated on the whole time horizon $T = 650$ and a rolling window with 240.
# Portfolio categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Portfolios</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>1-6</td>
</tr>
<tr>
<td>value interaction</td>
<td>7-9</td>
</tr>
<tr>
<td>momentum</td>
<td>10-13</td>
</tr>
<tr>
<td>reversal</td>
<td>14-18</td>
</tr>
<tr>
<td>investment</td>
<td>19-23</td>
</tr>
<tr>
<td>profitability</td>
<td>24-26</td>
</tr>
<tr>
<td>other</td>
<td>27-37</td>
</tr>
<tr>
<td>Portfolio</td>
<td>Mean</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Ind. Rel. Rev. (L.V.)</td>
<td>1.33</td>
</tr>
<tr>
<td>Industry Mom. Rev.</td>
<td>1.18</td>
</tr>
<tr>
<td>Industry Rel. Reversals</td>
<td>1.00</td>
</tr>
<tr>
<td>Seasonality</td>
<td>0.81</td>
</tr>
<tr>
<td>Value-Profitability</td>
<td>0.75</td>
</tr>
<tr>
<td>Momentum (12m)</td>
<td>1.28</td>
</tr>
<tr>
<td>Value-Mom-Prof.</td>
<td>0.84</td>
</tr>
<tr>
<td>Investment/Assets</td>
<td>0.48</td>
</tr>
<tr>
<td>Composite Issuance</td>
<td>0.45</td>
</tr>
<tr>
<td>Investment Growth</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Number of factors

Onatski (2010): Eigenvalue-ratio test

- **RP-PCA**: 6 factors
- **PCA**: 5 factors
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></th>
<th>Sharpe-Ratio</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>
</tr>
<tr>
<td>Size and BM</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>BM and Investment</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>Size and Accrual</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>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>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>
</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>
</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>
</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>
</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>
</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>
</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>
</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>
</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>
</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>
</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>
</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>
</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>
</tr>
</tbody>
</table>