Estimating Asset Pricing Factors from Large-Dimensional Data

Markus Pelger ¹  Martin Lettau ²

¹Stanford University
²UC Berkeley

April 21st, 2017

Machine Learning in Finance Workshop 2017
DSI Columbia University and Bloomberg LP
Motivation: Asset Pricing with Risk Factors

The Challenge of Asset Pricing

- Motivation: Asset Pricing with Risk Factors
- Goals of this paper: Bring order into „factor chaos“
- Summarize the pricing information of a large number of assets in a small number of factors

Fundamental Question: Which factors are really important in explaining expected returns? Which are subsumed by others?

Problem: „Chaos“ in asset pricing factors: Over 330 potential asset pricing factors published!

Fundamental Insight: Arbitrage Pricing Theory: Prices of financial assets should be explained by systematic risk factors.

Most Important Question in Finance: Why are prices different for different assets?
<table>
<thead>
<tr>
<th>Why is it important?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find underpriced assets and earn &quot;alpha&quot;</td>
</tr>
<tr>
<td>Arbitrage opportunities</td>
</tr>
<tr>
<td>Factors</td>
</tr>
<tr>
<td>Returns</td>
</tr>
<tr>
<td>Smart beta investments = investment strategies based on</td>
</tr>
<tr>
<td>Find investment decision with high risk-adjusted expected returns</td>
</tr>
<tr>
<td>Investment decisions</td>
</tr>
<tr>
<td>Factors allows to manage systematic risk exposure</td>
</tr>
<tr>
<td>Factors explain risk-return trade-off</td>
</tr>
<tr>
<td>Risk-management</td>
</tr>
<tr>
<td>Importance of finding the &quot;right&quot; factors</td>
</tr>
</tbody>
</table>
Contribution of this paper

- 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
3 times higher Sharpe-ratio than benchmark factors (PCA).

New factors much smaller pricing errors in- and out-of sample.

Empirical results:

PCA does not find all high Sharpe-ratio factors.

Component Analysis (PCA) dominates conventional approaches (Principal Investment).

Estimator discovers high Sharpe-ratio factors important for asset pricing.

No "blackbox approach". Asymptotic distribution theory for weak and strong factors.
Approximate Factor Model

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

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

Matrix notation

$$X = \underbrace{F_{t_1}^T \Lambda_1}_{T \times N} + \underbrace{e_{t_1}}_{T \times N}$$

- $N$ assets (large)
- $T$ time-series observation (large)
- $K$ systematic factors (fixed)
- $F$, $\Lambda$ and $e$ are unknown
Approximate Factor Model

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

\[ \text{Var}[F] \epsilon = [\epsilon']X \epsilon \]

explained by the risk-premium of the factors:

\[ \text{Var}(\epsilon) \text{ and Var}(F) \epsilon \text{uncorrelated} \]

Arbitrage-Pricing Theory (APT): The expected excess return is

\[ E[X_i] = E[F] \epsilon \]

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

\[ \text{Var}(\epsilon) \text{ and non-systematic risk (} F \text{ and } \epsilon \text{ uncorrelated}): \]

\[ \text{Var}(\epsilon) + \text{Var}(F) = \text{Var}(X) \]

Strong Factors

Weak Factors
The Model: Estimation of Latent Factors

- 
  \[ f = \text{estimator for factors}: \mathbf{f} = \mathbf{X} \mathbf{\lambda}_N \mathbf{1}^{-1} \]
  
  Eigenvectors of largest eigenvalues estimate loadings \( \mathbf{\lambda} \).

- 
  \[ \mathbb{X} = \mathbb{X} - \mathbb{X}_\perp \mathbf{1}^{-1} \]
  
  Eigenvectors of largest eigenvalues estimate loadings \( \mathbf{\lambda} \).

\[ \hat{\mathbf{X}} = \mathbf{X}_\perp \mathbf{1}^{-1} \]

- 
  Apply PCA to covariance matrix with overweighted mean

\[ \hat{\mathbf{F}} = \mathbf{X} \mathbf{\lambda}_N \mathbf{1}^{-1} \mathbf{1} \]

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

Conventional approach: PCA (Principal Component Analysis)
Model

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

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

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 \( \gamma \)
RP-PCA combines first and second moments efficiently.

Using more information leads to more efficient estimates.

Ignores first moment.

PCA of a covariance matrix uses only the second moment but

Information Interpretation (GMM Interpretation)

Penalizes low Sharpe-ratios.

Penalized PCA: Search for factors explaining the time-series but

Requires the time-series errors to be small as well.

Protects against spurious factors with vanishing loadings as it

Selects factors with small cross-sectional pricing errors (alpha’s).

Combines variation and pricing error criterion functions:

Interpretation of Risk-Premium-PCA (RP-PCA):
The Model

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

Signal-strengthening: Intuitively, the matrix $\text{Var}(\epsilon) + V \left( \frac{\text{Var}(F)}{\text{Var}(F)} + I \right) V$

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

$\text{var} = \text{var}(F) (1 + \gamma) \mu F + \text{Var}(\epsilon)

\text{with} \ X \rightarrow \Lambda^{-1} \Sigma \Lambda F + (1 + \gamma) \mu F$

$\text{converges to}

\text{Signal-strengthening:} \ X \rightarrow \Lambda^{-1} \Sigma \Lambda F + (1 + \gamma) \mu F$

$\text{with} \ X \rightarrow \Lambda^{-1} \Sigma \Lambda F + (1 + \gamma) \mu F$

$\text{converges to}

\text{Signal-strengthening:} \ X \rightarrow \Lambda^{-1} \Sigma \Lambda F + (1 + \gamma) \mu F$
Illustration: Anomaly-sorted portfolios (Size and accrual)
RP-PCA significantly better than PCA and quantile-sorted factors.

\[
\text{RMS} \alpha: \text{Root-mean-squared pricing errors} \quad \text{Pricing error } \alpha_i = E[X_i] - E[F_i] \Lambda_i \quad \text{Cross-sectional pricing errors } \alpha
\]

SR: Maximum Sharpe-ratio of linear combination of factors

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

<table>
<thead>
<tr>
<th>Specific</th>
<th>0.173</th>
<th>0.155</th>
<th>3.44</th>
<th>0.344</th>
<th>0.154</th>
<th>61.979</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fama-French</td>
<td>0.344</td>
<td>0.154</td>
<td>44.570</td>
<td>0.305</td>
<td>0.068</td>
<td>44.570</td>
</tr>
<tr>
<td>PCA</td>
<td>0.135</td>
<td>0.141</td>
<td>89.946</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP-PCA</td>
<td>0.135</td>
<td>0.141</td>
<td>89.946</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Portfolio Data: In-sample (Size and accrual)
Cross-sectional $\alpha$’s for sorted portfolios (Size and Accrual)

![Graph showing pricing errors for Size and Accrual]

⇒ RP-PCA avoids large pricing errors due to penalty term.
Loadings for statistical factors (Size and Accrual)

⇒ RP-PCA detects accrual factor while 3rd PCA factor is noise.
Maximal Incremental Sharpe Ratio

<table>
<thead>
<tr>
<th></th>
<th>PCA</th>
<th>RP-PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Factor</td>
<td>0.134</td>
<td>0.137</td>
</tr>
<tr>
<td>2 Factors</td>
<td>0.135</td>
<td>0.139</td>
</tr>
<tr>
<td>3 Factors</td>
<td>0.135</td>
<td>0.305</td>
</tr>
</tbody>
</table>

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

$\Rightarrow$ 1st and 2nd PCA and RP-PCA factors the same.

$\Rightarrow$ Better performance of RP-PCA because of third accrual factor.
Portfolio Data: Objective function (Size and Accrual)

<table>
<thead>
<tr>
<th></th>
<th>PCA TS</th>
<th>RP-PCA TS</th>
<th>PCA XS</th>
<th>RP-PCA XS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Factor</td>
<td>3.308</td>
<td>3.617</td>
<td>0.014</td>
<td>0.002</td>
</tr>
<tr>
<td>2 Factors</td>
<td>1.937</td>
<td>2.240</td>
<td>0.014</td>
<td>0.002</td>
</tr>
<tr>
<td>3 Factors</td>
<td>1.623</td>
<td>1.751</td>
<td>0.014</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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

⇒ RP-PCA and PCA explain the same amount of variation.
⇒ PR-PCA explains cross-sectional pricing much better.
⇒ Motivation for risk-premium weight $\gamma = 100$. 
# Portfolio Data: Out-of-sample (Size and Accrual)

<table>
<thead>
<tr>
<th></th>
<th>Out-of-sample</th>
<th>In-sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP-PCA</td>
<td>0.097</td>
<td>0.090</td>
</tr>
<tr>
<td>PCA</td>
<td>0.128</td>
<td>0.146</td>
</tr>
<tr>
<td>Fama-French 5</td>
<td>0.111</td>
<td>0.102</td>
</tr>
<tr>
<td>Specific</td>
<td>0.134</td>
<td>0.126</td>
</tr>
</tbody>
</table>

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

⇒ RP-PCA performs better in- and out-of-sample.
Weak vs. Strong Factor models

Optimal and relatively large \( \Rightarrow \)

PCA

RP-PCA detects weak factors which cannot be detected by

\( \epsilon \).\( \text{e.g. value factor} \)

Interpretation: weak factors affect a smaller fraction of assets,

Weak factor model \( (\Lambda \bot \Lambda_{\text{bounded}}) \)

Optimal and relatively small \( \Rightarrow \)

RP-PCA always more efficient than PCA

\( \epsilon \).\( \text{e.g. market factor} \)

Interpretation: strong factors affect most assets (proportional

Strong factor model \( (\Lambda \Vert N \bot N_{\text{bounded}}) \)

The Model

Weak vs. Strong Factors

Appendix

Empirical Results

Conclusion

Time-Variation

Strong Factors

Introduction
Strong vs. weak factor models

- Consequences for eigenvalues of $\frac{1}{T} X^T X$:

  - Strong factors lead to exploding eigenvalues
  - Weak factors lead to large but bounded eigenvalues

Empirical evidence (equity data): Strong and weak factors:

- Strong factors typically substantially larger than rest of spectrum
- Usually 10 times larger than the 2nd
- 1st eigenvalue typically substantially larger than rest of spectrum
- 2nd and 3rd eigenvalues typically stand out, but similar magnitudes as the rest of the spectrum

The Model

Strong vs. Strong Factors

Intro Model Illustration

Appendix

Strong Factors

Empirical Results

Conclusion

Time-Variation

Strong Factors

Weak Factors
Weak Factors

• The bulk, majority of eigenvalues
• The extremes, a few large outliers

Eigenvalues of sample covariance matrix separate into two areas:

In statistical model: spiked covariance models from random matrix

\( \Lambda_{\text{bounded}} \) (after normalizing factor variances)

\( V \) fraction of assets:

Weak factors either have a small variance or affect a smaller distribution (under certain conditions)

Bulk spectrum converges to Generalized Marchenko-Pastur
New tools necessary!

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

RP-PCA uses zero mean

Problem: All models in the literature assume that random processes have mean zero

Weak factor model with phase transition

Onatski (2012): Weak factor model with phase transition

Orthogonal to true eigenvectors if eigenvalues too small

Phase transition phenomena: estimated eigenvectors orthogonal to true eigenvectors if the true eigenvalue is below some critical threshold

To the bulk of the spectrum if the true eigenvalue is below the bulk spectrum

A biased value characterized by the Stieltjes transform of the

Large eigenvalues converge either to

Weak Factor Model
Strong Factors

Time-Variation

Empirical Results

Conclusion

Appendix

Intro

Model

Illustration

Weak Factors

Strong Factors

Time-Variation

Empirical Results

Conclusion

Appendix

Intro

Model

Illustration

Weak Factors

Signal: If \( \gamma \neq 0 \) and \( \gamma > 1 \) then RP-PCA signal always larger than PCA

\( \Lambda \Sigma F \Lambda > \theta_{\text{RP-PCA}} \)

K largest eigenvalues \( \Lambda \Sigma F \Lambda > \theta_{\text{RP-PCA}} \)

"Signal" matrix for RP-PCA:

\( \Lambda \Sigma F \Pi F \Lambda \)

K largest eigenvalues \( \Lambda \Sigma F \Pi F \Lambda \)

"Signal" matrix for PCA of covariance matrix:

\( \theta_{\text{PCA}} \)

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

\( \theta_{\text{RP-PCA}} \)
Theorem 1: \( \text{Risk-Premium PCA} \) strictly dominates PCA for \( \theta \) strictly increasing in \( \theta \).

Based on closed-form expression choose optimal RP-weight and are known in closed-form.

Critical value \( \theta_{\text{cr}} \) and function \( B(. \) depend only on the noise distribution.

\[
\begin{align*}
\theta_{\text{cr}} & < \theta \quad \text{if} \quad \theta_{\text{cr}} \quad \text{otherwise} \\
\end{align*}
\]

\[
\begin{align*}
\frac{0}{I + \theta_{\text{cr}} (1 + \theta_{\text{cr}} )^2} & \quad \text{if} \quad \theta_{\text{cr}} > \theta_{\text{crit}} \\
\end{align*}
\]

with

\[
\begin{align*}
\begin{pmatrix}
\rho_1 & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \vdots \\
0 & \cdots & \rho_d & 0 \\
0 & \cdots & 0 & 1
\end{pmatrix} & \quad \text{rotation} \\
\end{align*}
\]

The correlation of the estimated with the true factors converges to

\[
\text{with} \\
\begin{align*}
\text{corr}(F, F' \rho (\theta)) & \quad \text{rotation} \\
\end{align*}
\]
Strong Factors affect most assets: e.g. market factor

Statistical model: Bai and Ng (2002) and Bai (2003) framework

\[ \Lambda > \Lambda_{\text{bounded}}(\text{after normalizing factor variances}) \]

Assumptions essentially identical to Bai (2003)

RP-PCA provides a more efficient estimator of the loadings

Asymptotically normal distributed

Factors and loadings can be estimated consistently and are

Assumptions essentially identical to Bai (2003)
Asymptotic Efficiency

RP-PCA and PCA are both consistent

Optimal \( \gamma \) typically smaller than optimal value from weak factor model

RP-PCA (i.e. \( \gamma < -1 \)) always more efficient than PCA

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

Asymptotic Distribution (up to rotation)

Asymptotically \( \hat{F} \) behaves like OLS regression of \( A \) on \( X \).

Asymptotically \( \hat{\Lambda} \) behaves like OLS regression of \( \Lambda \) on \( X \).

Asymptotically \( \hat{\Lambda} \) behaves like OLS regression of \( F \) on \( X \).

Asymptotically \( \hat{\Lambda} \) behaves like OLS regression of \( FW \) on \( XW \) with \( W = \begin{pmatrix} I^T + \gamma \end{pmatrix} \), where \( I \) is a \( T \times 1 \) vector of 1's.

RP-PCA under slightly stronger assumptions as in Bai (2003):

PCA under assumptions of Bai (2003):

Asymptotically \( \hat{F} \) behaves like OLS regression of \( A \) on \( X \).

Asymptotically \( \hat{\Lambda} \) behaves like OLS regression of \( F \) on \( X \).

Optimal \( \gamma \) typically smaller than optimal value from weak factor model.

Asymptotic Efficiency

Choose RP-weight \( \gamma \) to obtain smallest asymptotic variance of estimators.
Time-varying loadings

Allow for general interactions between covariates
- We include the pricing error penalty
- Idea: Similar to projected PCA (Fan, Liao and Wang (2009)), but
  - Factors and loading function are latent
  - E.g. characteristics like size, book-to-market ratio, past returns,
  - Loadings are function of l covariates $Z_{i,t-1}, \ldots, Z_{i,t-l}$ with $l = 1, \ldots, l$

\[
X_{i,t} = \gamma_{i}^{\prime} \Gamma_{i}^{\prime} Z_{i,t-1} \ldots Z_{i,t-l} + e_{i,t} + \varepsilon_{i,t}
\]

Observe panel of excess returns and l covariates $Z_{i,t-1}^{l}$
Time-varying loadings

\[ \hat{\varphi}_k(Z_i, t - 1) = \sum_{m=1}^{M} b_m \phi_m(Z_i, t - 1) \]

Approximate nonlinear function \( g_k(\cdot) \) by sieve method

Apply RP-PCA to projected data

Intuition: Projection creates portfolios sorted on any functional conditional tree sorting projection

Obtain arbitrary interactions and break curse of dimensionality by

Appropriate basis functions (e.g. splines, kernels)
Empirical Results

Portfolio Data

- Monthly return data from July 1963 to December 2013
- 13 double sorted portfolios (consisting of 25 portfolios) from Kenneth French's website and 49 industry portfolios
- 606 double sorted portfolios

Data}

Factors

- Specific factors: market + two specific anomalies Long-short
- and investment

Fama-French factor model: market, size, value, profitability

RP-PCA: $K = 3$ and $\gamma = 100$

PCA: $K = 3$

Specific factors: market + two specific anomalies Long-short
### Empirical Results

#### Pricing errors $\alpha$ (in-sample)

<table>
<thead>
<tr>
<th>Factor Pairs</th>
<th>RP-PCA</th>
<th>PCA</th>
<th>FF 5</th>
<th>Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size and BM</td>
<td>0.13</td>
<td>0.14</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td><strong>BM and Investment</strong></td>
<td><strong>0.07</strong></td>
<td><strong>0.12</strong></td>
<td><strong>0.14</strong></td>
<td><strong>0.13</strong></td>
</tr>
<tr>
<td>BM and Operating Profits</td>
<td>0.11</td>
<td>0.12</td>
<td>0.14</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>Size and Accrual</strong></td>
<td><strong>0.07</strong></td>
<td><strong>0.14</strong></td>
<td><strong>0.15</strong></td>
<td><strong>0.16</strong></td>
</tr>
<tr>
<td>Size and Beta</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>Size and Investment</td>
<td>0.11</td>
<td>0.13</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>Size and Operating Profits</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>Size and Short-Term Reversal</td>
<td>0.15</td>
<td>0.16</td>
<td>0.24</td>
<td>0.33</td>
</tr>
<tr>
<td>Size and Long-Term Reversal</td>
<td>0.11</td>
<td>0.13</td>
<td>0.09</td>
<td>0.20</td>
</tr>
<tr>
<td>Size and Res. Var.</td>
<td>0.17</td>
<td>0.18</td>
<td>0.21</td>
<td>0.22</td>
</tr>
<tr>
<td>Size and Total Var.</td>
<td>0.18</td>
<td>0.19</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>Operating Profits and Investment</strong></td>
<td><strong>0.11</strong></td>
<td><strong>0.14</strong></td>
<td><strong>0.12</strong></td>
<td><strong>0.14</strong></td>
</tr>
<tr>
<td>Size and Net Share Iss.</td>
<td>0.14</td>
<td>0.16</td>
<td>0.13</td>
<td>0.17</td>
</tr>
<tr>
<td>49 Industries</td>
<td>0.14</td>
<td>0.16</td>
<td>0.13</td>
<td>0.29</td>
</tr>
</tbody>
</table>
### Empirical Results

**Pricing errors $\alpha$ (out-of-sample)**

<table>
<thead>
<tr>
<th></th>
<th>RP-PCA</th>
<th>PCA</th>
<th>FF 5</th>
<th>Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size and BM</td>
<td>0.17</td>
<td>0.19</td>
<td>0.14</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>BM and Investment</strong></td>
<td>0.12</td>
<td>0.16</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td>BM and Operating Profits</td>
<td>0.15</td>
<td>0.18</td>
<td>0.15</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>Size and Accrual</strong></td>
<td>0.10</td>
<td>0.13</td>
<td>0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>Size and Beta</td>
<td>0.09</td>
<td>0.10</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>Size and Investment</td>
<td>0.14</td>
<td>0.17</td>
<td>0.12</td>
<td>0.19</td>
</tr>
<tr>
<td>Size and Operating Profits</td>
<td>0.09</td>
<td>0.12</td>
<td>0.09</td>
<td>0.18</td>
</tr>
<tr>
<td>Size and Short-Term Reversal</td>
<td>0.17</td>
<td>0.19</td>
<td>0.09</td>
<td>0.18</td>
</tr>
<tr>
<td>Size and Long-Term Reversal</td>
<td>0.13</td>
<td>0.14</td>
<td>0.09</td>
<td>0.14</td>
</tr>
<tr>
<td>Size and Res. Var.</td>
<td>0.17</td>
<td>0.20</td>
<td>0.18</td>
<td>0.26</td>
</tr>
<tr>
<td>Size and Total Var.</td>
<td>0.17</td>
<td>0.21</td>
<td>0.20</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>Operating Profits and Investment</strong></td>
<td>0.13</td>
<td>0.17</td>
<td>0.13</td>
<td>0.16</td>
</tr>
<tr>
<td>Size and Net Share Iss.</td>
<td>0.14</td>
<td>0.21</td>
<td>0.16</td>
<td>0.18</td>
</tr>
<tr>
<td>49 Industries</td>
<td>0.26</td>
<td>0.24</td>
<td>0.21</td>
<td>0.25</td>
</tr>
</tbody>
</table>
## Empirical Results

### Maximum Sharpe-Ratios

<table>
<thead>
<tr>
<th>Factor Combination</th>
<th>RP-PCA</th>
<th>PCA</th>
<th>Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size and BM</td>
<td>0.25</td>
<td>0.22</td>
<td>0.16</td>
</tr>
<tr>
<td>BM and Investment</td>
<td><strong>0.26</strong></td>
<td><strong>0.17</strong></td>
<td><strong>0.24</strong></td>
</tr>
<tr>
<td>BM and Operating Profits</td>
<td>0.24</td>
<td>0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>Size and Accrual</td>
<td><strong>0.30</strong></td>
<td><strong>0.13</strong></td>
<td><strong>0.17</strong></td>
</tr>
<tr>
<td>Size and Beta</td>
<td>0.23</td>
<td>0.23</td>
<td>0.17</td>
</tr>
<tr>
<td>Size and Investment</td>
<td>0.30</td>
<td>0.26</td>
<td>0.23</td>
</tr>
<tr>
<td>Size and Operating Profits</td>
<td>0.22</td>
<td>0.21</td>
<td>0.18</td>
</tr>
<tr>
<td>Size and Short-Term Reversal</td>
<td>0.26</td>
<td>0.20</td>
<td>0.25</td>
</tr>
<tr>
<td>Size and Long-Term Reversal</td>
<td>0.23</td>
<td>0.18</td>
<td>0.15</td>
</tr>
<tr>
<td>Size and Res. Var.</td>
<td>0.33</td>
<td>0.30</td>
<td>0.32</td>
</tr>
<tr>
<td>Size and Total Var.</td>
<td>0.32</td>
<td>0.28</td>
<td>0.32</td>
</tr>
<tr>
<td>Operating Profits and Investment</td>
<td><strong>0.31</strong></td>
<td><strong>0.24</strong></td>
<td><strong>0.34</strong></td>
</tr>
<tr>
<td>Size and Net Share Iss.</td>
<td><strong>0.33</strong></td>
<td><strong>0.25</strong></td>
<td><strong>0.35</strong></td>
</tr>
<tr>
<td>49 Industries</td>
<td>0.35</td>
<td>0.25</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Empirical Results

Portfolio Data

Portfolio Data

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

Volatility factors:
Specific: Market, value, value-momentum-profitability and
from Kenneth French’s website.

Fama-French 5: The five factor model of Fama-French

49 industry portfolios from Kenneth French’s website

Santosh (2015)

Novy-Marx and Velikov (2014) data: 150 portfolios sorted
according to 15 anomalies (same data as in Kozak, Nagel and
for $N = 199$ portfolios

Monthly return data from July 1963 to December 2013 ($T = 606$)
Empirical Results

Portfolio Data: 15 Novy-Mark factors and portfolios

Beta Arbitrage
Long Run Reversal
Momentum
Idiosyncratic Vol
ValMom
ValMomProf
Piotrofki F-Score
Investment
Asset Growth
Net Issuance
Accruals
Value Prof
Value
Gross Profitability
Size
Specific factors (Market, Value, Value-Momentum, Profitability)

RP-PCA strongly dominates PCA and Fama-French 5 factors

Portfolio Data: In-sample

<table>
<thead>
<tr>
<th></th>
<th>Fama-MacBeth</th>
<th>Specific</th>
<th>Fama-French</th>
<th>PCA</th>
<th>RP-PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal Sharpe-Ratios</td>
<td>731.392</td>
<td>0.413</td>
<td>0.344</td>
<td>0.155</td>
<td>0.417</td>
</tr>
<tr>
<td>Root-mean-squared pricing errors and risk-premium</td>
<td>801.013</td>
<td>0.225</td>
<td>0.213</td>
<td>0.135</td>
<td>0.417</td>
</tr>
<tr>
<td>820.804</td>
<td>820.804</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>729.944</td>
<td>729.944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Weight γ = 100.
Portfolio Data: Out-of-sample

<table>
<thead>
<tr>
<th></th>
<th>Out-of-sample</th>
<th>In-sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP-PCA</td>
<td>0.178</td>
<td>0.145</td>
</tr>
<tr>
<td>PCA</td>
<td>0.202</td>
<td>0.208</td>
</tr>
<tr>
<td>Fama-French 5</td>
<td>0.182</td>
<td>0.182</td>
</tr>
<tr>
<td>Specific</td>
<td>0.154</td>
<td>0.137</td>
</tr>
</tbody>
</table>

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

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

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. Invertible linear transformations. Table: Generalized Correlations between specific factors and statistical factors.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>0.997</td>
<td>0.032</td>
<td>0.888</td>
<td>0.925</td>
<td>0.967</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Portfolio Data: Interpreting factors.
### Methodology

**New estimator for estimating priced latent factors from large data sets**

- Combines time-series and cross-sectional criterion function sets
- Asymptotic theory under weak and strong factor model assumption
- Detects weak factors with high Sharpe-ratio
- More efficient than conventional PCA

### Empirical Results

- Strongly dominates estimation based on PCA of the covariance matrix
- Potential to provide benchmark factors for horse races
- Promising empirical results

**Promising empirical results**
Literature (partial list)

- Bryzgalova (2016): Spurious factors
- Koizumi, Nagel and Santoshi (2015): PCA based factors
- Clarke (2015): Level, slope and curvature for stocks
- Harvey and Liu (2015): Lucky factors

Asset-picking factors

- Bai (2003): Distribution theory

Random matrices

- Benachoucha-Georges and Nadakuditi (2011): Perturbation of Large
- Pelger (2016), Ait-Sahalia and Xiu (2015): High-frequency
- Fan et al. (2013): Sparse matrices in factor modeling
- Bai and Ng (2002): Determining the number of factors

- Bai (2003): Distribution theory

- A1

Large-dimensional factor models with weak factors

- Large-dimenional factor models with strong factors

- Large-dimensional factor models with weak factors (based on random matrix theory)

- Large-dimensional factor models with strong factors
Motivation for PCA

$\text{A minimizes time-series objective function}$

$\forall X \in \mathbb{R}^{N \times T}$ 

$A \propto \text{eigen vectors of the first } K \text{ largest eigenvalues of } \text{VAR}(X)$

$\forall X = V F - \mathbb{E}[X] = \mathbb{E}[\text{VAR}(X)]$

$\text{Projection matrix } M = V V^T - I$

$\text{Error (non-systematic risk): } e = X - F$

$\text{Time-series objective function:}$

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

$\text{Minimize the unexplained variance:}$
Spurious factor detection (Bryzgalova (2016))

Factors not identified

Why not estimate factors with cross-sectional objective function?

\[
\frac{1}{T} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\left( \frac{1}{T} \sum_{t=1}^{T} X_i \right) - \frac{1}{T} \sum_{t=1}^{T} \hat{F}_i \Lambda}{\Lambda^T} \right)^2 \geq \frac{T}{I} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\left( \frac{1}{T} \sum_{t=1}^{T} X_i \right) - \frac{1}{T} \sum_{t=1}^{T} \hat{F}_i \Lambda}{\Lambda^T} \right)^2
\]

\[
\frac{1}{T} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\left( \frac{1}{T} \sum_{t=1}^{T} X_i \right) - \frac{1}{T} \sum_{t=1}^{T} \hat{F}_i \Lambda}{\Lambda^T} \right)^2 \geq \frac{T}{I} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\left( \frac{1}{T} \sum_{t=1}^{T} X_i \right) - \frac{1}{T} \sum_{t=1}^{T} \hat{F}_i \Lambda}{\Lambda^T} \right)^2
\]

Minimize cross-sectional expected pricing error:

Cross-sectional objective function:

\[
\min 
\frac{1}{T} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\left( \frac{1}{T} \sum_{t=1}^{T} X_i \right) - \frac{1}{T} \sum_{t=1}^{T} \hat{F}_i \Lambda}{\Lambda^T} \right)^2
\]
The objective function is minimized by the eigenvectors of the largest eigenvalues of the first $K$.

Estimator for the common component $\mathbf{C} = \mathbf{F} \mathbf{A}$ is $\hat{\mathbf{C}} = \hat{\mathbf{F}} \hat{\mathbf{A}}$.

Estimator for factors: $\hat{\mathbf{F}} = \mathbf{X} \left( \mathbf{V} \mathbf{V}^{\top} \right)^{-1} \mathbf{V} \mathbf{V}^{\top} \mathbf{X}$.

Estimator for loadings: proportional to eigenvectors of the first $K$.

Combined objective function: Risk-Premium-PCA
The Model

Optimal choice of $\Omega^*$: GLS type argument

$\Omega^* \equiv -\Omega$ corresponds to $\text{PCA}$ of a correlation matrix.

Today: Only $\Omega$ equal to inverse of a diagonal matrix of standard deviations.

Factors and loadings can be estimated by applying $\text{PCA}$ to cross-sectional weighting matrix $\Omega$.

Cross-sectional weighting matrix $\Omega$:

$$
(\Omega_{XX}^{-1} + I)^{-1} X \Omega
$$

Straightforward extension to weighted objective function:

$$
\min_{\Lambda} \text{trace } M \Lambda Q > X I + \gamma X Q \Lambda
$$

s.t.

$$
(\Omega_{XX}^{-1} + I) (\Omega_{XX}^{-1} + I)^{-1} X \Omega
$$

Weighted Combined objective function:
Assume that $c > 0$ with $\frac{1}{N} \rightarrow \infty$. Then

$$\mathcal{V} = \mathcal{V} \perp \mathcal{V}$$

independent of $e$ and $F$ (e.g. $\mathcal{V} \perp \mathcal{V} \mathcal{V} \perp \mathcal{V}$

The column vectors of the loadings $\mathbf{V}$ are orthogonally invariant and have bounded first two moments. The factors $F$ are uncorrelated among each other and are independent of $e$ and $\Lambda$ and have bounded first two moments.

The supremum of the support is $b$. The non-random spectral distribution function with compact support. The empirical eigenvalue distribution function of $\mathcal{V} \perp \mathcal{V}$ converges to a non-random spectral distribution function with compact support. The residual matrix can be represented as $e = \mathcal{V} \epsilon$ with $\epsilon \sim \mathcal{N}(0, I)$. The

Assumption 1: Weak Factor Model
Definition: Weak Factor Model

\[
\left( \left( e \perp \frac{I}{N} \right) - \left( \frac{1}{N} \right) \right) \quad \text{trace} \quad \frac{N}{\mathbb{C}} \quad \lim \quad \text{as} \quad \frac{1}{N} \quad \Rightarrow \quad \frac{1}{N} \quad \lim \quad \text{as} \quad \frac{1}{N} \quad \Rightarrow \quad (z) \quad \mathcal{B}
\]

\[
\left( \frac{1}{N} \right) \quad \text{trace} \quad \frac{N}{\mathbb{C}} \quad \lim \quad \text{as} \quad \frac{1}{N} \quad \Rightarrow \quad \frac{1}{N} \quad \lim \quad \text{as} \quad \frac{1}{N} \quad \Rightarrow \quad (z) \quad \mathcal{C}
\]

\[
\left( \left( e \perp \frac{I}{N} \right) - \left( \frac{1}{N} \right) \right) \quad \text{trace} \quad \frac{N}{\mathbb{C}} \quad \lim \quad \text{as} \quad \frac{1}{N} \quad \Rightarrow \quad \frac{1}{N} \quad \lim \quad \text{as} \quad \frac{1}{N} \quad \Rightarrow \quad (z) \quad \mathcal{D}
\]

The almost sure limit of the Cauchy transform of the ordered eigenvalues is denoted by \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \). The average idiosyncratic noise is:

\[
\frac{N}{(\Xi)} \quad \text{trace} \quad \frac{N}{\mathbb{C}}
\]
**Weak Factor Model**

**Intro**

**Model Illustration**

**Weak Factors**

**Strong Factors**

**Time-Variation**

**Empirical Results**

**Conclusion**

**Appendix**

**Estimator**

**PCA special case of RP-PCA**

\[
S^{-1}\hat{S} = S^{-1} X (X' \tilde{X} + \tilde{X}' X) S^{-1} = S^{-1} X (X' \tilde{X} + \tilde{X}' X) S^{-1}
\]

**Intuition**: Largest $K$ "true" eigenvalues of $S^{-1}$.

\[
\begin{pmatrix}
\frac{\phi^2}{\sigma^2} + \frac{\phi^2}{\sigma^2} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \frac{\phi^2}{\sigma^2} + \frac{\phi^2}{\sigma^2}
\end{pmatrix}
\]

\[
M_{VAR} = \frac{\phi^2}{\sigma^2} \mathbf{I} + \frac{\phi^2}{\sigma^2} \mathbf{I}
\]

**Signal Matrix for Covariance PCA**

\[
\Sigma F + c \sigma^2 e I_K = \begin{pmatrix}
\sigma^2 F_1 + c \sigma^2 e & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \sigma^2 F_K + c \sigma^2 e
\end{pmatrix}
\]
Lemma: Covariance PCA

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

$$\lambda_i \overset{p}{\to} \begin{cases} G^{-1} \left( \frac{1}{\sigma^2_{F_i} + c\sigma^2_e} \right) \left( \begin{array}{c} 1 \\ \vdots \\ 1 \\ \sigma^2_{\text{crit}} \\ \vdots \\ \sigma^2_{\text{crit}} \end{array} \right) & \text{if } \sigma^2_{F_i} + c\sigma^2_e > \sigma^2_{\text{crit}} \\ b & \text{otherwise} \end{cases}$$

The correlation between the estimated and true factors converges to

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

with

$$\rho_i^2 \overset{p}{\to} \begin{cases} 1 & \text{if } \sigma^2_{F_i} + c\sigma^2_e > \sigma^2_{\text{crit}} \\ \frac{1}{1 + (\sigma^2_{F_i} + c\sigma^2_e)B(\lambda_i)} & \text{otherwise} \end{cases}$$
**Weak Factor Model**

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

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

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

The correlation between the estimated and true factors converges to

\[
\widehat{\text{Corr}}(F, \hat{F}) \xrightarrow{p} \begin{pmatrix} q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & q_K \end{pmatrix}
\]

with

\[
q_i^2 \xrightarrow{p} \begin{cases} 
1 - \frac{c\sigma^4_e}{\sigma^4_{F_i}} & \text{if } \sigma^2_{F_i} + c\sigma^2_e > \sigma^2_{\text{crit}} \\
\frac{1 - \frac{c\sigma^4_e}{\sigma^4_{F_i}}}{1 + \frac{c\sigma^2_e}{\sigma^2_{F_i}} + \frac{\sigma^4_e}{\sigma^4_{F_i}} (c^2 - c)} & \text{otherwise}
\end{cases}
\]
Define $\tilde{\gamma} = \sqrt{\gamma + 1} - 1$ and note that $(1 + \tilde{\gamma})^2 = 1 + \gamma$.

Intuition: $\theta_1, \ldots, \theta_{K+1}$ largest $K+1$ "true" eigenvalues of $S^\gamma$.

Let $U = \begin{pmatrix} \theta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_{K+1} \end{pmatrix}$

Denote by $U$ the corresponding orthonormal eigenvectors.

Projection on demeaned factors and on mean operator:

$M_{RP} = \begin{pmatrix} \theta_1 \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \null$
**Theorem 1: Risk-Premium PCA under weak factor model**

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

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

The correlation of the estimated with the true factors converges to

$$\hat{\text{Corr}}(F, \hat{F}) \overset{p}{\rightarrow} (I_K \ 0) \ \check{U}$$

with

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

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

\[D_K = \text{diag} \left( \begin{pmatrix} \hat{\theta}_1 & \cdots & \hat{\theta}_K \end{pmatrix} \right)\]
Weak Factor Model

Lemma: Detection of weak factors

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

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

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

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

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

⇒ For $\mu_F \neq 0$ RP-PCA always better than PCA.
Example: One-factor model

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

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

and has the eigenvalues:

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

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

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

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

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

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

\[
\hat{\text{Corr}}(F, \hat{F}) \xrightarrow{p} \frac{\rho_1}{\sqrt{\rho_1^2 + (1 - \rho_1^2) \left(\frac{\theta_1 - (\sigma_F^2 + c\sigma_e^2))^2 + 1}{\mu^2\sigma_F^2(1+\gamma)}\right)}}
\]
Strong Factors affect most assets: e.g. market factor

Assumptions essentially identical to Bai (2003)

RP-PCA provides a more efficient estimator of the loadings normal distributed

Factors and loadings can be estimated consistently and are asymptotically statistically model: Bai and Ng (2002) and Bai (2003) framework

\[ V \frac{N}{N} \]

\( N \) bounded (after normalizing factor variances)
Asymptotic expansions (under slightly stronger assumptions as in Bai (2003)):  

\[
\sqrt{T} \left[ \begin{pmatrix} I \end{pmatrix}^{d_o} + \left( \frac{1}{N^\infty} \right)^{d_o} O + \frac{1}{\sqrt{N}} \mathcal{V} \frac{\mathcal{V}^T}{T} (V \perp \mathcal{V}) \right] = \left( \frac{1}{\sqrt{T}} F \right)^T \mathcal{V} \frac{\mathcal{V}^T}{T} (V \perp \mathcal{V}) \mathcal{V} \frac{\mathcal{V}^T}{T} F \right] \mathcal{N}^{\infty} \\
\sqrt{N} \left[ \begin{pmatrix} I \end{pmatrix}^{d_o} + \left( \frac{N}{T^\infty} \right)^{d_o} O + \frac{W}{\sqrt{T}} \mathcal{W} \frac{\mathcal{W}^T}{T} (W \perp \mathcal{W}) \right] = \left( \frac{1}{\sqrt{N}} \Lambda \right)^T \mathcal{V} \frac{\mathcal{V}^T}{T} (V \perp \mathcal{V}) \mathcal{V} \frac{\mathcal{V}^T}{T} \Lambda \right] \mathcal{N}^{\infty}
\]

Asymptotically \( \hat{\Lambda} \) behaves like OLS regression of \( F \) on \( X \).

Asymptotically \( \hat{F} \) behaves like OLS regression of \( \Lambda \) on \( X \).

An empirical illustration of the model.
Assumption 2: Strong Factor Model

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

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_t e_{t,i} \xrightarrow{D} N(0, \Omega) \quad \Omega = \begin{pmatrix} \Omega_{1,1} & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{pmatrix}
\]
Assumption 2 holds and \( \gamma \in [-1, \infty) \). Then:

\[
(\Phi, 0)_N \overset{\mathcal{D}}{\rightarrow} \left( \begin{array}{c}
\gamma \\
\gamma \mathcal{N}
\end{array} \right) \begin{array}{c}
\mathcal{N} \\
\mathcal{D}
\end{array}
\]

If \( \frac{1}{N} \) then \( 0 \rightarrow \frac{1}{N} \). Then 

For any choice of \( \gamma \), the factors, loadings and common components can be estimated consistently pointwise.

The asymptotic distribution of the common component depends on \( \gamma \).

The asymptotic distribution of the factors is not affected by the choice of \( \gamma \) if only if does not go to zero. For

\[ 0 \leftarrow \frac{N}{T} \]
Strong Factor Model

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

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

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

with

$$\Omega = \sigma^2_e \left( \frac{\sigma^2_F + \mu^2_F (1 + \gamma)^2}{\sigma^2_F + \mu^2_F (1 + \gamma)} \right)^2$$

$\Rightarrow$ Optimal choice minimizing the asymptotic variance is risk-premium weight $\gamma = 0$.

$\Rightarrow$ Choosing $\gamma = -1$, i.e. the covariance matrix for factor estimation, is not efficient.
Simulation parameters:

Factors: $K = 4$

$N = 250$ and $T = 350$. 

Errors: Cross-sectional and time-series correlation and heteroskedasticity in the residuals. Half of the variation due to non-systematic risk.

Loadings normalized such that $\Lambda_i^\prime \Lambda_i > 1$ and $\Lambda_i^\prime = 1$ for $i = 2, 3, 4$.

$N(0, 4, 0.4^2)$; Sharp-ratio of 0.4.

4. Factor has a small variance but high Sharp-ratio. It follows $N(0, 0.4, 1^2)$; Sharp-ratio of 0.4.

3. Factor follows $N(0, 0.4, 1^2)$; Sharp-ratio of 0.4.

2. Factor represents an industry factors following $N(0.1, 1^2)$; Sharp-ratio of 0.4.

1. Factor represents the market with $N(1.2, 9); Sharp-ratio of 0.15$.
Around half of the variation is due to systematic risk and half is non-systematic.

Simulation parameters:
- Signal-to-noise ratio: \( \sigma^2 = 10 \)
- Empirical results: following \( N(1, 0.2) \)
- Time-series heteroskedasticity: \( \sigma^2 \) is a diagonal matrix with independent elements.
- Time-series correlation in errors: \( \rho = 0.7 \)
- Cross-sectional correlation in errors: \( \rho \) on the right four off-diagonals with \( \rho = 0.7 \)
- Cross-sectional heteroskedasticity: \( \sigma^2_{\text{N}} \) is a Toeplitz matrix with \( \beta^2 = 0.7 \)
- Time-series correlation in errors: \( \rho \) creates an AR(1) model with \( \rho = 0.7 \)
- \( N \times T \) matrix and follows a multivariate standard normal distribution
- Residuals are modeled as \( \epsilon = \sigma^2 \sigma^2_{\text{T}} \times \text{AR}(1) \times \text{N} \times T \)
Figure: Sample path of the first four factors and the estimated factor processes. $\gamma = 50$. 
Simulation

<table>
<thead>
<tr>
<th></th>
<th>PCA Var</th>
<th>PCA Corr</th>
<th>RP-PCA</th>
<th>RP-PCA Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Factor</td>
<td>0.094</td>
<td>0.086</td>
<td>0.042</td>
<td>0.040</td>
</tr>
<tr>
<td>2. Factor</td>
<td>0.023</td>
<td>0.022</td>
<td>0.025</td>
<td>0.022</td>
</tr>
<tr>
<td>3. Factor</td>
<td>0.100</td>
<td>0.095</td>
<td>0.079</td>
<td>0.074</td>
</tr>
<tr>
<td><strong>4. Factor</strong></td>
<td><strong>0.312</strong></td>
<td><strong>0.312</strong></td>
<td><strong>0.183</strong></td>
<td><strong>0.170</strong></td>
</tr>
</tbody>
</table>

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

A 25
Simulation

Squared correlations between estimated and true factor based on the weak factor model prediction and Monte-Carlo simulations for different variances of the factor. The Sharpe-ratio of the factor is 1, i.e. the mean equals $\mu_F = \sigma_F$. The normalized variance of the factors is $\sigma_F^2 \cdot N$. 
Figure: 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 $\sigma_e^2 = 1$ and $c = 1$. For the correlated residuals we assume that $\Sigma^{1/2}$ is a Toeplitz matrix with $\beta, \beta, \beta, \beta^2$ on the right four off-diagonals with $\beta = 0.7$. 
Simulation

<table>
<thead>
<tr>
<th>True Factors</th>
<th>PCA Var</th>
<th>PCA Corr</th>
<th>RP-PCA</th>
<th>PR-PCA Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR</td>
<td>1.330</td>
<td>0.515</td>
<td>0.517</td>
<td><strong>0.865</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>PCA Var</th>
<th>PCA Corr</th>
<th>RP-PCA</th>
<th>RP-PCA Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Factor</td>
<td>1.20</td>
<td>1.10</td>
<td>1.11</td>
<td>1.16</td>
<td>1.16</td>
</tr>
<tr>
<td>2. Factor</td>
<td>0.10</td>
<td>0.11</td>
<td>0.10</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>3. Factor</td>
<td>0.40</td>
<td>0.31</td>
<td>0.31</td>
<td>0.49</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>4. Factor</strong></td>
<td><strong>0.40</strong></td>
<td><strong>0.08</strong></td>
<td><strong>0.08</strong></td>
<td><strong>0.21</strong></td>
<td><strong>0.22</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>PCA Var</th>
<th>PCA Corr</th>
<th>RP-PCA</th>
<th>RP-PCA Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Factor</td>
<td>9.000</td>
<td>8.608</td>
<td>8.615</td>
<td>8.494</td>
<td>8.510</td>
</tr>
<tr>
<td>2. Factor</td>
<td>1.000</td>
<td>0.697</td>
<td>0.716</td>
<td>0.683</td>
<td>0.706</td>
</tr>
<tr>
<td>3. Factor</td>
<td>1.000</td>
<td>0.801</td>
<td>0.820</td>
<td>0.674</td>
<td>0.690</td>
</tr>
<tr>
<td>4. Factor</td>
<td>0.160</td>
<td>0.028</td>
<td>0.028</td>
<td>0.066</td>
<td>0.070</td>
</tr>
</tbody>
</table>

**Table:** Estimated variance of factors. $\gamma = 50$. 
Cross-sectional $\alpha$’s out-of-sample (Size and Accrual)

⇒ RP-PCA avoids large pricing errors due to penalty term.
Predicted excess return in-sample (Size and Accrual)
Predicted excess return out-of-sample (Size and Accrual)

- OLS out-of-sample RP-PCA
- OLS out-of-sample PCA
- OLS out-of-sample 5 Fama-French factors
- OLS out-of-sample Specific factors
### Fama-MacBeth Test-Statistics: $\chi^2_{22}: 34(95 \%)$

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

- OLS out-of-sample RP-PCA
- OLS out-of-sample PCA
- OLS out-of-sample 5 Fama-French factors
- OLS out-of-sample Specific factors
Maximal Incremental Sharpe Ratio

<table>
<thead>
<tr>
<th></th>
<th>PCA</th>
<th>RP-PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Factor</td>
<td>0.127</td>
<td>0.137</td>
</tr>
<tr>
<td>2 Factors</td>
<td>0.149</td>
<td><strong>0.381</strong></td>
</tr>
<tr>
<td>3 Factors</td>
<td>0.153</td>
<td><strong>0.412</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>PCA TS</th>
<th>RP-PCA TS</th>
<th>PCA XS</th>
<th>RP-PCA XS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Factor</td>
<td>44.771</td>
<td>51.623</td>
<td>0.298</td>
<td>0.037</td>
</tr>
<tr>
<td>2 Factors</td>
<td>39.846</td>
<td>42.326</td>
<td>0.268</td>
<td>0.001</td>
</tr>
<tr>
<td>3 Factors</td>
<td>36.112</td>
<td>37.849</td>
<td>0.263</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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

⇒ RP-PCA and PCA explain the same amount of variation.
⇒ PR-PCA explains cross-sectional pricing much better.
⇒ Motivation for risk-premium weight $\gamma = 100$. 
Table: Maximal Sharpe-ratios, root-mean-squared pricing errors and Fama-MacBeth test statistics for different set of factors. For the statistical factor estimators we use $K = 3$ factors and $\gamma = 100$. 

<table>
<thead>
<tr>
<th>Factor</th>
<th>SR</th>
<th>RMS $\alpha$</th>
<th>Fama-MacBeth</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP-PCA</td>
<td>0.256</td>
<td>0.074</td>
<td>14.520</td>
</tr>
<tr>
<td>PCA</td>
<td>0.169</td>
<td>0.123</td>
<td>37.038</td>
</tr>
<tr>
<td>Fama-French</td>
<td>0.344</td>
<td>0.140</td>
<td>26.144</td>
</tr>
<tr>
<td>Specific</td>
<td>0.236</td>
<td>0.127</td>
<td>31.611</td>
</tr>
</tbody>
</table>
Cross-sectional $\alpha$’s for sorted portfolios (BM and Investment)
Loadings for statistical factors (BM and Investment)
Maximal Incremental Sharpe Ratio (BM and Investment)

<table>
<thead>
<tr>
<th></th>
<th>PCA</th>
<th>RP-PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Factor</td>
<td>0.144</td>
<td>0.149</td>
</tr>
<tr>
<td>2 Factors</td>
<td>0.167</td>
<td>0.193</td>
</tr>
<tr>
<td>3 Factors</td>
<td>0.169</td>
<td>0.256</td>
</tr>
</tbody>
</table>

**Table:** Maximal Sharpe-ratio by adding factors incrementally. For the statistical factor estimators we use $K = 3$ factors and $\gamma = 100$. 
Portfolio Data: Objective function (BM and Investment)

<table>
<thead>
<tr>
<th></th>
<th>PCA TS</th>
<th>RP-PCA TS</th>
<th>PCA XS</th>
<th>RP-PCA XS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Factor</td>
<td>5.543</td>
<td>5.989</td>
<td>0.021</td>
<td>0.002</td>
</tr>
<tr>
<td>2 Factors</td>
<td>4.416</td>
<td>4.647</td>
<td>0.014</td>
<td>0.001</td>
</tr>
<tr>
<td>3 Factors</td>
<td>3.944</td>
<td>4.098</td>
<td>0.013</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Table:** Time-series and cross-sectional objective functions.
### Portfolio Data: Out-of-sample (BM and Investment)

<table>
<thead>
<tr>
<th></th>
<th>Out-of-sample</th>
<th>In-sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP-PCA</td>
<td>0.123</td>
<td>0.065</td>
</tr>
<tr>
<td>PCA</td>
<td>0.157</td>
<td>0.156</td>
</tr>
<tr>
<td>Fama-French 5</td>
<td>0.111</td>
<td>0.103</td>
</tr>
<tr>
<td>Specific</td>
<td>0.138</td>
<td>0.138</td>
</tr>
</tbody>
</table>

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

---

A 43
Cross-sectional $\alpha$’s out-of-sample (BM and Investment)
Predicted excess return in-sample (BM and Investment)
Predicted excess return out-of-sample (BM and Invest.)