Interpretable Proximate Factors for Large Dimensions

Markus Pelger \(^1\)  \quad Ruoxuan Xiong \(^2\)

Stanford University

2018 SoFiE Summer School in Chicago on Machine Learning and Empirical Asset Pricing
July 27th, 2018
Motivation: What are the factors?

Statistical Factor Analysis

- Factor models are widely used in big data settings
  - Reduce data dimensionality
  - Factors are traded extensively
  - Problem: Which factors should be used?
- Statistical (latent) factors perform well
  - Factors estimated from principle component analysis (PCA)
  - Weighted averages of all assets/portfolios
  - Problem: Hard to interpret

Goals of this paper:

Create interpretable proximate factors

- Shrink most small assets’ weights to zero to get proximate factors
  ⇒ More interpretable
  ⇒ Significantly lower transaction costs when trading factors
Motivation

Contribution of this paper

**Contribution**

- **This Paper:** Estimation of interpretable proximate 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. Proximate factors approximate latent factors very well with a few assets without sparse structure in population loadings
  4. Only 5-10% of the cross-sectional observations with the largest exposure are needed for proximate factors
Theoretical Results

- Asymptotic probabilistic lower bound for generalized correlations of proximate factors with population factors
- Guidance on how to construct proximate factors

Empirical Results

- Very good approximation to population factors with 5-10% portfolios, measured by generalized correlation, variance explained, pricing error and Sharpe-ratio
- Interpret statistical latent factors for
  - Double-sorted portfolio data
  - 370 single-sorted anomaly portfolios
  - 128 macroeconomic variables
Literature (partial list)

- Large-dimensional factor models with PCA
  - Bai and Ng (2002): Number of factors
  - Bai (2003): Distribution theory
  - Fan et al. (2013): Sparse matrices in factor modeling
  - Fan et al. (2016): Projected PCA for time-varying loadings
  - Pelger (2017), Aït-Sahalia and Xiu (2015): High-frequency
  - Kelly, Pruitt and Su (2017): IPCA

- Factor models with penalty term
  - Bai and Ng (2017): Robust PCA with ridge shrinkage
  - Lettau and Pelger (2018): Risk-Premium PCA with pricing penalty
  - Zhou et al. (2006): Sparse PCA (low dimension)
Empirical Application: Size and Value Portfolios

- 25 portfolios formed on size and value (07/1963-10/2017, 3 factors, daily data)

(a) Generalized Correlation

(b) Variance Explained

(c) RMS Pricing Error

(d) Max Sharpe Ratio
Empirical Application: Size and Book-to-market Portfolios

Figure: Portfolio weights of 1. statistical factor

⇒ Equally weighted market factor
Empirical Application: Size and Book-to-market Portfolios

Figure: Portfolio weights of 2. statistical factor

- Small-minus-big size factor
- Proximate factor with 4 largest weights correlation 0.88 with size factor
Empirical Application: Size and Book-to-market Portfolios

Figure: Portfolio weights of 3. statistical factor

⇒ High-minus-low value factor
⇒ Proximate factor with 4 largest weights correlation 0.91 with value factor
The Model

Approximate Factor Model

- Observe panel data of $N$ assets over $T$ time periods:

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

- Matrix notation

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

- $N$ 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):

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

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

\[ \Rightarrow \] Idiosyncratic risk can be weakly correlated

Estimation: PCA (Principal Component Analysis)

- Apply PCA to the sample covariance matrix: $\frac{1}{T}XX^\top - \bar{X}\bar{X}^\top$ with $\bar{X}$ = sample mean of asset excess returns

- Eigenvectors of largest eigenvalues estimate loadings $\hat{\Lambda}$

- $\hat{F}$ estimator for factors: $\hat{F} = X\hat{\Lambda}^\top (\hat{\Lambda}^\top\hat{\Lambda})^{-1} = \frac{1}{N}X^\top\hat{\Lambda}$
Method to Construct Proximate Factors

Proximate Factors

- Sparse loadings $\tilde{\Lambda}$ are obtained from
  - Select finitely many $m$ loadings with largest absolute value from $\hat{\Lambda}_k$ for all $k$
  - Shrink estimated loadings $\hat{\Lambda}$ to 0 except for $m$ largest values
  - Divide by column norms, i.e. $\tilde{\lambda}_k^T \tilde{\lambda}_k = 1$

- Proximate factors $\tilde{F} = X^T \tilde{\Lambda} (\tilde{\Lambda}^T \tilde{\Lambda})^{-1}$
Closeness between Proximate Factors and Latent Factors

Closeness measure

- For 1-factor model: Correlation between $\tilde{F}$ and $F$.
- Problem for multiple factors: Factors are only identified up to invertible linear transformations $\Rightarrow$ Need measure for closeness between span of two vector spaces.
- For multi-factor model: The ”closeness” between $\tilde{F}$ and $F$ is measured by generalized correlation:
  
  Total generalized correlation measure:
  
  $$
  \rho = \text{trace} \left( (F^T F / T)^{-1} (F^T \tilde{F} / T) (\tilde{F}^T \tilde{F} / T)^{-1} (\tilde{F}^T F / T) \right)
  $$

  - $\rho = 0$: $\tilde{F}$ and $F$ are orthogonal
  - $\rho = K$: $\tilde{F}$ and $F$ span the same space
Intuition: Why does picking largest elements in $\hat{\Lambda}$ work?

- Consider one factor and one nonzero element in $\tilde{\Lambda}$:
  
  $F = [f_{1t}] \in \mathbb{R}^{T \times 1}$, $\Lambda = [\lambda_{1,i}] \in \mathbb{R}^{N \times 1}$

  $\tilde{\Lambda} = [\tilde{\lambda}_{1,i}]$ is sparse. Assume nonzero element in $\tilde{\lambda}_{1,i}$ is $\tilde{\lambda}_{1,1} = 1$, so $\tilde{\Lambda}^T \tilde{\Lambda} = I$.

  \[
  \tilde{F} = X^T \tilde{\Lambda} = F \Lambda^T \tilde{\Lambda} + e^T \tilde{\Lambda} = f_{1} \lambda_{1,1} + e_1
  \]

- Assume

  \[
  f_{1,t} \sim (0, \sigma_f^2), \quad e_{1,t} \overset{iid}{\sim} (0, \sigma_e^2)
  \]

  \[
  \frac{f_{1}^T f_{1}}{T} \rightarrow \sigma_f^2, \quad \frac{e_{1}^T e_{1}}{T} \rightarrow \sigma_e^2
  \]

- Define signal-to-noise ratio $s = \frac{\sigma_f}{\sigma_e}$
Intuition: Why pick the largest elements in $\hat{\Lambda}$?

$$\rho = \text{tr} \left( (F^T F / T)^{-1} (F^T \tilde{F} / T) (\tilde{F}^T \tilde{F} / T)^{-1} (\tilde{F}^T F / T) \right)$$

$$= \left( \frac{f_1^T (f_1 \lambda_{1,1} + e_1) / T}{(f_1^T f_1 / T)^{1/2}} \left( (f_1 \lambda_{1,1} + e_1)^T (f_1 \lambda_{1,1} + e_1) / T \right)^{1/2} \right)^2$$

$$\rightarrow \frac{\lambda_{1,1}^2}{\lambda_{1,1}^2 + 1/s^2}$$

- (Generalized) correlation increases in size of loading $|\lambda_{1,1}|$.
- (Generalized) correlation increases in signal-to-noise ratio $s$.
- No sparsity in population loadings assumed!
Asymptotic results

- Proximate factors $\tilde{F}$ are in general not consistent.
  - Consider one-factor model
    $\tilde{F} = X^T\tilde{\Lambda} = F\Lambda^T\tilde{\Lambda} + e^T\tilde{\Lambda}$
  - Idiosyncratic component not diversified away
  - Assume $e_{i,t} \overset{iid}{\sim} (0, \sigma_{e.,t}^2)$, then each element in $e^T\tilde{\Lambda}$ has
    $$\text{Var} \left( \sum_{i=1}^{m} \tilde{\lambda}_{1,1,i} e_{1,i,t} \right) = \sum_{i=1}^{m} \tilde{\lambda}_{1,1,i}^2 \sigma_{e.,t}^2 = \sigma_{e.,t}^2 \rightarrow 0$$

- Instead we provide probabilistic lower bound for (generalized) correlation $\rho$ given a target correlation level $\rho_0$:
  $$P(\rho > \rho_0)$$
Assumptions

Similar assumptions as in Bai and Ng (2002)

1. **Factors:** $E \|f_t\|^4 \leq M < \infty$ and $\frac{1}{T} \sum_{t=1}^{T} f_t f_t^T \overset{P}{\to} \Sigma_F$ for some $K \times K$ positive definite matrix $\Sigma_F = \text{diag}(\sigma_{f1}^2, \sigma_{f2}^2, \cdots, \sigma_{fr}^2)$.

2. **Loadings:** Random variables $\max_i \|\lambda_{j,i}\| = O_p(1)$ and $\Lambda^\top \Lambda / N \to \Sigma_\Lambda$, independent of factors and errors.

3. **Systematic factors:** Eigenvalues of $\Sigma_\Lambda \Sigma_F$ bounded away from 0 and $\infty$.

4. **Residuals:** Weak Dependency
   - Bounded eigenvalues and sparsity of $\Sigma_e$
   - $e$ weakly dependent with $F$
   - Light tails

$\Rightarrow$ Uniform convergence result for loadings $\forall i, \exists H$,

$$\max_{i \leq N} \|\hat{\lambda}_{(i)} - H\lambda_{(i)}\| = O_p \left(\frac{1}{\sqrt{N}} + \frac{N^{1/4}}{\sqrt{T}}\right).$$
One factor case

**Theorem 1: Probabilistic lower bound**

Assume $K = 1$ factor and population loadings $\lambda_{1,i}$ are i.i.d for all $i$. For any $\rho_0$ we have for $N, T \to \infty$

$$P(\rho > \rho_0) \geq 1 - \sum_{j=0}^{m-1} \binom{N}{j} (1 - F_{|\lambda_{1,i}|}(y_m))^j F_{|\lambda_{1,i}|}(y_m)^{N-j} + o_p(1)$$

where $y_m = \sqrt{\frac{1+f(m)}{m} \frac{\sigma_e^2}{\sigma_{f_1}^2} \frac{\rho_0}{1-\rho_0}}$ and $F_{|\lambda_{1,i}|}(y) = P(|\lambda_{1,i}| \leq y)$. $f(m)$ measures the maximum errors’ total correlations among any $m$ cross-section units.
One factor case

- Denote the lower probability bound for \( P(\rho > \rho_0) \) by
  \[
  \underline{p} = 1 - \sum_{j=0}^{m-1} \binom{N}{j} (1 - F_{|\lambda_{1,i}|}(y_m))^j F_{|\lambda_{1,i}|}(y_m)^{N-j}
  \]
- It holds:
  \[
  \frac{\partial \underline{p}}{\partial F_{|\lambda_{1,i}|}(y_m)} < 0
  \]
- \( \underline{p} \) is decreasing in \( F_{|\lambda_{1,i}|}(y_m) \). Hence \( \underline{p} \) is
  - increasing in \( s = \sigma_{f_1}/\sigma_e \)
  - increasing in the dispersion of the distribution of \( |\lambda_{1,i}| \)
  - increasing in \( m \) for i.i.d errors

Corollary 1

Under the assumptions of Theorem 1 if \( \lambda_i \) have unbounded support then for any fixed \( \rho_0 < 1 \)

\[
P(\rho > \rho_0) \rightarrow 1
\]
Theorem 2: Distribution of correlation

Assume: \( K = 1 \) factor and there exists sequences of constants \( \{a_{1,N} > 0\} \) and \( \{b_{1,N}\} \) such that

\[
P\left( \left| \lambda_{1,(1)} - b_{1,N} \right| / a_{1,N} \leq z \right) \rightarrow G_1(z),
\]

Then for \( N, T \rightarrow \infty \)

\[
P \left( \rho \geq \frac{\sigma_f^2 (a_{1,N}z + b_{1,N})^2}{(1 + f(m))\sigma_e^2 + \sigma_f^2 (a_{1,N}z + b_{1,N})^2} \right) \geq 1 - G_{1,m}(z) + o_p(1),
\]

\( G_1 \) is the Generalized Extreme Value (GEV) distribution function,

\[
G_1 = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}
\]
One factor case: Extreme value theory

A few examples for $G_1$ and $a_{1,N}$ and $b_{1,N}$ for $\lambda_{1,i}$:

1. $G_1 \sim$ Gumbel distribution:
   - Standard normal distribution ($\lambda_i \sim N(0,1)$): $a_{1,N} = \frac{1}{N\phi(b_{1,N})}$ and $b_{1,N} = \Phi^{-1}(1 - 1/N)$, where $\phi(\cdot), \Phi(\cdot)$ are pdf and cdf of standard normal.
   - Exponential distribution ($\lambda_i \sim \exp(1)$): $a_{1,N} = 1, b_{1,N} = N$

2. $G_1 \sim$ Frechet distribution:
   - $F_\lambda(x) = \exp(-1/x)$: $a_{1,N} = N, b_{1,N} = 0$.

3. $G_1 \sim$ Weibull distribution:
   - Uniform: distribution ($\lambda_i \sim Uniform(0,1)$): $a_{1,N} = 1/N, b_{1,N} = 1$. 
Theorem 3: Multiple Factor: Simple Case with GEV

Assume each row in $V = \Lambda H$ has only one nonzero value and there exists sequences $\{a_{j,N} > 0\}$ and $\{b_{j,N}\}$ s.t.

$$P\left(\left( v_{j,(1)} - b_{j,N} \right)/a_{j,N} \leq z \right) \to G_j(z)$$

for all $j$. Then for $N, T \to \infty$:

$$P \left( \rho \geq K - \frac{(1+f(m))\sigma^2}{m} \sum_{j=1}^{K} \frac{1}{s_j(a_{j,N}z + b_{j,N})^2} \right) \geq \prod_{j=1}^{K} (1 - G_{j,m}^*(z)) + o_P(1),$$

where $G_{j,m}$ is the Generalized Extreme Value (GEV) distribution function.

⇒ Theory is relaxed to more than one nonzero values in each row.
Multi Factor Case

Multiple Factors

**Multiple Factor: Rotate and threshold**

- Assume there exists orthonormal matrix $P$ s.t. large values in columns of $W^P = \Lambda HSP$ do not overlap (almost orthogonal)

- $m$ nonzero entries in $\tilde{W}_j$ are the largest in $\hat{W}_j$ satisfying $\max_{j,k \neq j} |\hat{w}_{i,k}^P / \hat{w}_{i,j}^P| < c$ and are standardized by

$$\tilde{W}^P = \begin{bmatrix} \frac{\hat{w}_1^P \odot M_1}{\| \hat{w}_1^P \odot M_1 \|} & \frac{\hat{w}_2^P \odot M_2}{\| \hat{w}_2^P \odot M_2 \|} & \cdots & \frac{\hat{w}_K^P \odot M_K}{\| \hat{w}_K^P \odot M_K \|} \end{bmatrix}.$$  

- The proximate factors are

$$\tilde{F}^P = X^T \tilde{W}^P ((\tilde{W}^P)^T \tilde{W}^P)^{-1} = X^T \tilde{W}^P$$

- Generalized Correlation

$$\rho = tr \left( (F^T F / T)^{-1} (F^T \tilde{F}^P / T)( (\tilde{F}^P)^T \tilde{F}^P / T)^{-1} ((\tilde{F}^P)^T F / T) \right)$$
Multi Factor Case

Multiple Factors

**Theorem 4: Rotate and threshold**

Let $\tilde{w}^P_{(m),j}$ be the $m$-th order statistic of the entries in $|w^P_j|$ that satisfy
\[
\max_{j,k \neq j} |w^P_{i,k}/w^P_{i,j}| < c \quad \text{and assume that the cumulative density function of} \quad \tilde{w}^P_{(m),j} \quad \text{is continuous. Then for a particular threshold} \quad 0 < \rho_0 < K \quad \text{and a fixed} \quad m, \quad \text{we have}
\]
\[
\lim_{N,T \to \infty} P(\rho > \rho_0) \geq \lim_{N \to \infty} P \left( \sum_{j=1}^{K} \frac{1}{(\tilde{w}^P_{(m),j})^2} < \frac{m(1 - \gamma)(K - \rho_0)}{(1 + f(m))\sigma_e^2} \right),
\]

where $\gamma = c(2 + c(K - 2))(K(K - 1))^{1/2}$. 
Denote the lower probability bound for $P(\rho > \rho_0)$ by $p$

It holds (very similar to the one factor case) that $p$ is

- increasing in $s_j = \sigma_{f_j}/\sigma_e$
- increasing in the dispersion of the distribution of $|\lambda_{j,i}|$
- increasing in $m$ for i.i.d errors
Empirical Results

Single-sorted Portfolios

**Portfolio Data**

- Monthly return data from 07/1963 to 12/2016 ($T = 638$) for $N = 370$ portfolios

- Kozak, Nagel and Santosh (2017) data: 370 decile portfolios sorted according to 37 anomaly characteristics, such as momentum, volatility, turnover, size and volume.
Single-sorted Portfolios

(a) Generalized Correlation  
(b) Variance Explained

Figure: Financial single-sorted portfolios: Generalized correlation between \( \tilde{F} \) and \( \hat{F} \) normalized by \( K \) and proportion of variance explained by \( \tilde{F} \) and \( \hat{F} \) as a function of non-zero loading elements \( m \), where \( K \) varies from 3 to 7. \((N = 370, \: T = 638)\)
## Single-sorted Portfolios

<table>
<thead>
<tr>
<th>m</th>
<th>$\hat{F}_1$</th>
<th>$\hat{F}_2$</th>
<th>$\hat{F}_3$</th>
<th>$\hat{F}_4$</th>
<th>$\hat{F}_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.993</td>
<td>0.992</td>
<td>0.771</td>
<td>0.918</td>
<td>0.837</td>
</tr>
<tr>
<td>20</td>
<td>0.995</td>
<td>0.948</td>
<td>0.883</td>
<td>0.949</td>
<td>0.890</td>
</tr>
<tr>
<td>30</td>
<td>0.996</td>
<td>0.965</td>
<td>0.935</td>
<td>0.966</td>
<td>0.910</td>
</tr>
<tr>
<td>40</td>
<td>0.997</td>
<td>0.971</td>
<td>0.958</td>
<td>0.975</td>
<td>0.923</td>
</tr>
</tbody>
</table>

**Table:** Financial single-sorted portfolios: Generalized correlation between each $\hat{F}_j$ and all $\tilde{F}$ for $K = 5$. These generalized correlations correspond to $R^2$ from a regression of each $\hat{F}_j$ on all $\tilde{F}$. 
Single-sorted Portfolios: Fourth Factor

Hard to interpret...

**Figure:** Financial single-sorted portfolios: Portfolio weights of 4th PCA factor.
The fourth factor is a momentum factor.

**Figure:** Financial single-sorted portfolios: Portfolio weights of 4th proximate factor. The sparse loading has 30 nonzero entries.
Macroeconomic data

Monthly data from 01/1959 to 02/2018 for \( N = 128 \) variables:

1. output and income
2. labor market
3. housing
4. consumption, orders and inventories
5. money and credit
6. interest and exchange rates
7. prices
8. stock market
Empirical Results

Macroeconomic data

(a) Generalized Correlation

(b) Variance Explained

Figure: Macroeconomic data: Generalized correlation between $\tilde{F}$ and $\hat{F}$ normalized by $K$ and proportion of variance explained by $\tilde{F}$ and $\hat{F}$, where $K$ varies from 4 to 20. ($N = 128$, $T = 707$)
### Empirical Results

#### Macroeconomic data

<table>
<thead>
<tr>
<th>m</th>
<th>$\hat{F}_1$</th>
<th>$\hat{F}_2$</th>
<th>$\hat{F}_3$</th>
<th>$\hat{F}_4$</th>
<th>$\hat{F}_5$</th>
<th>$\hat{F}_6$</th>
<th>$\hat{F}_7$</th>
<th>$\hat{F}_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.953</td>
<td>0.959</td>
<td>0.949</td>
<td>0.953</td>
<td>0.961</td>
<td>0.799</td>
<td>0.833</td>
<td>0.767</td>
</tr>
<tr>
<td>15</td>
<td>0.967</td>
<td>0.970</td>
<td>0.958</td>
<td>0.956</td>
<td>0.964</td>
<td>0.857</td>
<td>0.867</td>
<td>0.837</td>
</tr>
<tr>
<td>20</td>
<td>0.977</td>
<td>0.974</td>
<td>0.957</td>
<td>0.963</td>
<td>0.961</td>
<td>0.905</td>
<td>0.919</td>
<td>0.891</td>
</tr>
<tr>
<td>25</td>
<td>0.983</td>
<td>0.980</td>
<td>0.961</td>
<td>0.979</td>
<td>0.973</td>
<td>0.937</td>
<td>0.943</td>
<td>0.929</td>
</tr>
</tbody>
</table>

**Table:** Macroeconomic data: Generalized correlation between each $\hat{F}_j$ and $\tilde{F}$, where $K = 8$. These generalized correlations correspond to $R^2$ from a regression of each $\hat{F}_j$ on all $\tilde{F}$. 
Empirical Results

Macroeconomic data

Figure: Macroeconomic data: 8-factor model, each sparse loading in $\tilde{\Lambda}$ has 10 nonzero entries. Values in this figure represent the number of nonzero entries in a particular group for a particular sparse loading. The 8 groups are: 1. output and income; 2. labor market; 3. housing; 4. consumption, orders and inventories; 5. money and credit; 6. interest and exchange rates; 7. prices; 8. stock market.
Conclusion

Methodology

- Proximate factors (portfolios of a few assets) for latent population factors (portfolios of all assets)
- Simple thresholding estimator based on largest loadings
- Proximate factors approximate population factors well without sparsity assumption
- Asymptotic probabilistic lower bound for (generalized) correlation

⇒ Few observations summarize most of the information

Empirical Results

- Good approximation to population factors with 5-10% portfolios
Multiple Factors

Multiple Factor: Rotate and threshold

- Assume there exists orthonormal matrix $P$ s.t. large values in columns of $W^P = \Lambda HSP$ do not overlap (almost orthogonal)

- $m$ nonzero entries in $\tilde{W}_j$ are the largest in $\hat{W}_j$ satisfying $\max_{j,k \neq j} |\hat{w}_{i,k}/\hat{w}_{i,j}| < c$ and are standardized by

$$\tilde{W}^P = \left[ \frac{\hat{W}_1^P \odot M_1}{||\hat{W}_1^P \odot M_1||}, \frac{\hat{W}_2^P \odot M_2}{||\hat{W}_2^P \odot M_2||}, \ldots, \frac{\hat{W}_K^P \odot M_K}{||\hat{W}_K^P \odot M_K||} \right].$$

- The proximate factors are

$$\tilde{F}^P = X^T \tilde{W}^P ((\tilde{W}^P)^T \tilde{W}^P)^{-1} = X^T \tilde{W}^P$$

- Generalized Correlation

$$\rho = tr \left( (F^T F / T)^{-1} (F^T \tilde{F}^P / T) ((\tilde{F}^P)^T \tilde{F}^P / T)^{-1} ((\tilde{F}^P)^T F / T) \right)$$
Theorem 4: Rotate and threshold

Let \( \bar{w}^P_{(m),j} \) be the \( m \)-th order statistic of the entries in \( |w^P_j| \) that satisfy
\[
\max_{j,k \neq j} |w^P_{i,k}/w^P_{i,j}| < c
\]
and assume that the cumulative density function of \( \bar{w}^P_{(m),j} \) is continuous. Then for a particular threshold \( 0 < \rho_0 < K \) and a fixed \( m \), we have
\[
\lim_{N,T \to \infty} P(\rho > \rho_0) \geq \lim_{N \to \infty} P \left( \sum_{j=1}^{K} \frac{1}{(\bar{w}^P_{(m),j})^2} < \frac{m(1 - \gamma)(K - \rho_0)}{(1 + f(m))\sigma_e^2} \right),
\]
where \( \gamma = c(2 + c(K - 2))(K(K - 1))^{1/2} \).
Relationship with Lasso

Alternative approach with Lasso:

1. Estimate factors by PCA, i.e. \( X^T X \hat{F} = \hat{F} V \) with \( V \) matrix of eigenvalues.

2. Estimate loadings by minimizing \( \| X - \Lambda \hat{F}^T \|_F^2 + \alpha \| \Lambda \|_1 \).

   Divide the minimizer by its column norm (standardize each loading) to obtain \( \bar{\Lambda} \)

3. Proximate factors from Lasso approach are \( \bar{F} = X^T \bar{\Lambda} (\bar{\Lambda}^T \bar{\Lambda})^{-1} \)

⇒ Same selection of non-zero elements (for one factor case) but different weighting

⇒ Under certain conditions worse performance than thresholding approach

• Tuning parameter less transparent
Simulation

- Compare probabilistic lower bounds with Monte-Carlo simulations
- **Factors:** $K = 1$ or $K = 2$ and $F_t \sim N(0, \sigma_f^2)$
- **Loadings:** $\lambda_i \sim N(0, 1)$ i.i.d.
- **Residuals:** $\sigma_e = 1$ and $e_{t,i} \sim N(0, 1)$ i.i.d.
- Vary signal-to-noise ratio with $\sigma_f \in \{0.8, 1.0, 1.2\}$
- $N = 100$) and $T \in \{50, 100, 200\}$
- We analyze:
  - Probabilistic lower bound for $\rho_0 = 0.95$
  - Distribution of lower bound with extreme value distribution
Simulation: One factor with very strong signal

Figure: Probabilistic lower bound: $\sigma_f = 1.2, \rho_0 = 0.95$
Simulation: One factor with weaker signal

Figure: Probabilistic lower bound: $\sigma_f = 1.0, \rho_0 = 0.95$
Simulation: One factor with weak signal

Figure: Probabilistic lower bound: $\sigma_f = 0.8$, $\rho_0 = 0.95$
Simulation: One factor with increasing $N$

(a) One-factor model  
$\sigma_f = 1.0$

(b) Multi-factor model  
$\sigma_f = [1.2, 1.0]$

Figure: Probabilistic lower bound: $\rho_0 = 0.95$
Simulation: Two Factors

(a) $\sigma_f = [1.0, 0.8]$,   \hspace{1cm}  (b) $\sigma_f = [1.2, 1.0]$   \hspace{1cm}  (c) $\sigma_f = [1.5, 1.2]$

Figure: Probabilistic lower bound: $\rho_0 = 1.9$. 
Empirical Application: Size and Investment Portfolios

- 25 portfolios formed on size and investment (07/1963-10/2017, 3 factors, daily data)

(a) Generalized correlation
(b) Variance explained
(c) RMS pricing error
(d) Max Sharpe Ratio
Empirical Application: Size and Investment Portfolios

Figure: Portfolio weights of 1. statistical factor

⇒ Equally weighted market factor
Empirical Application: Size and Investment Portfolios

**Figure:** Portfolio weights of 2. statistical factor

⇒ Small-minus-big size factor

⇒ Proximate factor with 4 largest weights correlation 0.97 with size factor
Empirical Application: Size and Investment Portfolios

Figure: Portfolio weights of 3. statistical factor

⇒ High-minus-low value factor
⇒ Proximate factor with 4 largest weights correlation 0.79 with investment factor
### Single-sorted portfolios

<table>
<thead>
<tr>
<th>Anomaly characteristics</th>
<th>Anomaly characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Accruals - accrual</td>
<td>20 Momentum (12m) - mom12</td>
</tr>
<tr>
<td>2 Asset Turnover - aturnover</td>
<td>21 Momentum-Reversals - momrev</td>
</tr>
<tr>
<td>3 Cash Flows/Price - cfp</td>
<td>22 Net Operating Assets - noa</td>
</tr>
<tr>
<td>4 Composite Issuance - ciss</td>
<td>23 Price - price</td>
</tr>
<tr>
<td>5 Dividend/Price - divp</td>
<td>24 Gross Profitability - prof</td>
</tr>
<tr>
<td>6 Earnings/Price - ep</td>
<td>25 Return on Assets (A) - roaa</td>
</tr>
<tr>
<td>7 Gross Margins - gmargins</td>
<td>26 Return on Book Equity (A) - roea</td>
</tr>
<tr>
<td>8 Asset Growth - growth</td>
<td>27 Seasonality - season</td>
</tr>
<tr>
<td>9 Investment Growth - igrowth</td>
<td>28 Sales Growth - sgrowth</td>
</tr>
<tr>
<td>10 Industry Momentum - indmom</td>
<td>29 Share Volume - shvol</td>
</tr>
<tr>
<td>11 Industry Mom. Reversals - indmomrev</td>
<td>30 Size - size</td>
</tr>
<tr>
<td>12 Industry Rel. Reversals - indrev</td>
<td>31 Sales/Price - sp</td>
</tr>
<tr>
<td>13 Industry Rel. Rev. (L.V.) - indrevlv</td>
<td>32 Short-Term Reversals - strev</td>
</tr>
<tr>
<td>14 Investment/Assets - inv</td>
<td>33 Value-Momentum - valmom</td>
</tr>
<tr>
<td>15 Investment/Capital - invcap</td>
<td>34 Value-Momentum-Prof. - valmomprof</td>
</tr>
<tr>
<td>16 Idiosyncratic Volatility - ivol</td>
<td>35 Value-Profability - valprof</td>
</tr>
<tr>
<td>17 Leverage - lev</td>
<td>36 Value (A) - value</td>
</tr>
<tr>
<td>18 Long Run Reversals - Irrev</td>
<td>37 Value (M) - valuem</td>
</tr>
<tr>
<td>19 Momentum (6m) - mom</td>
<td></td>
</tr>
</tbody>
</table>
Macroeconomic data

(a) Generalized Correlation
(b) Variance Explained

Figure: Macroeconomic data: Generalized correlation between $\tilde{F}$ and $\hat{F}$ normalized by $K$ and proportion of variance explained by $\tilde{F}$ and $\hat{F}$, where $K$ varies from 4 to 20. ($N = 128, \ T = 707$)
Macroeconomic data

<table>
<thead>
<tr>
<th>m</th>
<th>$\hat{F}_1$</th>
<th>$\hat{F}_2$</th>
<th>$\hat{F}_3$</th>
<th>$\hat{F}_4$</th>
<th>$\hat{F}_5$</th>
<th>$\hat{F}_6$</th>
<th>$\hat{F}_7$</th>
<th>$\hat{F}_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.953</td>
<td>0.959</td>
<td>0.949</td>
<td>0.953</td>
<td>0.961</td>
<td>0.799</td>
<td>0.833</td>
<td>0.767</td>
</tr>
<tr>
<td>15</td>
<td>0.967</td>
<td>0.970</td>
<td>0.958</td>
<td>0.956</td>
<td>0.964</td>
<td>0.857</td>
<td>0.867</td>
<td>0.837</td>
</tr>
<tr>
<td>20</td>
<td>0.977</td>
<td>0.974</td>
<td>0.957</td>
<td>0.963</td>
<td>0.961</td>
<td>0.905</td>
<td>0.919</td>
<td>0.891</td>
</tr>
<tr>
<td>25</td>
<td>0.983</td>
<td>0.980</td>
<td>0.961</td>
<td>0.979</td>
<td>0.973</td>
<td>0.937</td>
<td>0.943</td>
<td>0.929</td>
</tr>
</tbody>
</table>

**Table:** Macroeconomic data: Generalized correlation between each $\hat{F}_j$ and $\tilde{F}$, where $K = 8$. These generalized correlations correspond to $R^2$ from a regression of each $\hat{F}_j$ on all $\tilde{F}$. 
Macroeconomic data

**Figure:** Macroeconomic data: 8-factor model, each sparse loading in $\tilde{\Lambda}$ has 10 nonzero entries. Values in this figure represent the number of nonzero entries in a particular group for a particular sparse loading. The 8 groups are: 1. output and income; 2. labor market; 3. housing; 4. consumption, orders and inventories; 5. money and credit; 6. interest and exchange rates; 7. prices; 8. stock market