Statistical Foundations of Data Science

Jianqing Fan

Princeton University

https://fan.princeton.edu

ZOOM ID Lectures: **970 4936 8998** Office Hours: **996 4030 7631**

Annotated Lecture Notes: web view

9. Covariance Learning and Factor Models

9.1. Principal Component Analysis

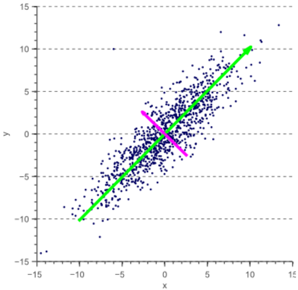
Principal Components

Data: $\{\mathbf{X}_i\}_{i=1}^n$ with an estimated covariance matrix $\widehat{\boldsymbol{\Sigma}}$. e.g. sample cov.

Principal Component: $\widehat{\xi}_1$ maximizes variance of projections $\{\mathbf{b}^T\mathbf{X}_i\}_{i=1}^n$:

$$\widehat{\xi}_1 = \operatorname{argmax}_{\|\mathbf{b}\|=1} \mathbf{b}^T \widehat{\boldsymbol{\Sigma}} \mathbf{b}.$$

★Sign of $\widehat{\xi}_1$ can not be determined uniquely.



2^{nd} PC: Subject to orthogal. (uncorrelatedness), maximize var. of projections $\{\mathbf{b}^T\mathbf{X}_i\}_{i=1}^n$.

$$\widehat{\xi}_2 = \operatorname{argmax}_{\mathbf{b}^T\widehat{\xi}_1=0, \|\mathbf{b}\|=1} \mathbf{b}^T \widehat{\boldsymbol{\Sigma}} \mathbf{b}.$$

Principal Components and Eigen-decomposition

Spectral decompositions: For any symmetric matrix,

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\lambda}_1 \widehat{\xi}_1 \widehat{\xi}_1^T + \cdots + \widehat{\lambda}_p \widehat{\xi}_p \widehat{\xi}_p^T,$$

★$\{\widehat{\xi}_j\}_{j=1}^p$ othonormal, i.e. $\widehat{\xi}_j^T \widehat{\xi}_k = 0$, for $i \neq j$, $\quad \|\widehat{\xi}_j\| = 1$;

★computed using R function "`eigen`".

Principal Components: Arrange $\{\widehat{\lambda}_j\}_{j=1}^p$ decreasingly. Then, ***k*th PC =** $\widehat{\xi}_k^T \mathbf{X}$ with $\mathrm{cov}(\widehat{\xi}_j^T \mathbf{X}, \widehat{\xi}_k^T \mathbf{X}) = \widehat{\lambda}_k I(j = k)$.

★$\widehat{\xi}_k$ is called k^{th} **direction**.

Two useful perturbation theorems

■Let $\widehat{\boldsymbol{\Sigma}}$ estimate $\boldsymbol{\Sigma}$, whose eigenpairs are $\{\lambda_i, \xi_i\}_{i=1}^p$

Weyl Theorem: $|\widehat{\lambda}_i - \lambda_i| \le \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_2$.

sin θ Theorem (Davis and Kahan, 1970): With $\widehat{\lambda}_0 = \infty$,

$$\|\widehat{\xi}_i - \xi_i\| \le \frac{\sqrt{2}\|\widehat{\Sigma} - \Sigma\|}{\min(|\widehat{\lambda}_{i-1} - \lambda_i|, |\lambda_i - \widehat{\lambda}_{i+1}|)}$$

Generalized to ★eigenspace ★rectangular (Wedin, 72) ★ℓ_∞-norm or ℓ_p-norm.

For elementary proofs, See Chen, Y., Chi, Y., Fan, J., Ma, C. (2020).

Spectral Methods for Data Science: A Statistical Perspective (Chap 2)

9.2. Covariance Learning and Factor Models

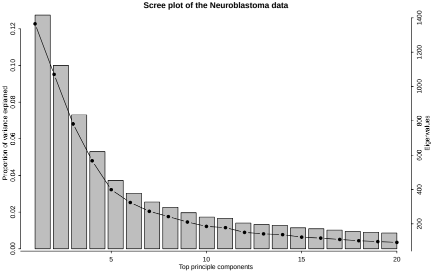
Dependence is a stylized feature in high-dim

Example 1: Neuroblastoma Data

★ 246 German patients diagnosed between 1989 and 2004, aged from 0 to 296 months (median 15 months).

★ Predict '3-y Event Free Survival", (56 "+" and 190 "−") using gene expressions $p = 10,707$

★ Predictors are correlated as shown in scree plot



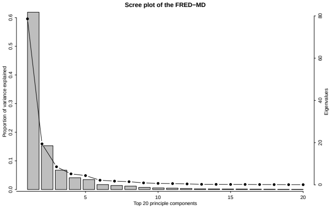
What factors drive the dependence?

Example 2: Forecasting Bond Risk Premia

★ $Y_t^{(n)}$: bond risk premia w/ maturity of $n = 2, 3, 4, 5$ yrs.

★ $\mathbf{X}_t$: 131 macroeconomic monthly series 1964-2018

(*Stock & Watson, 02,10; M. McCracken's web*).



★ $\mathbf{W}_t$: 8 macroeconomic(*e.g. NBER, 08; Stock, Watson 10*).

■ $W_1 =$ Linear comb. of five forward rates; ■ $W_2 =$ GDP; ■ $W_3 =$ Real category devel. index (CDI); ■ $W_4 =$Non-agriculture employment; ■ $W_5 =$ Real industrial production; ■ $W_6 =$ Real manufacturing and trade sales; ■ $W_7 =$ Real personal income less transfer; ■ $W_8 =$ Consumer price index (CPI)

How to use additional covariates?

Factor models

Factor model: $\mathbf{X}$ is p-dim variables, $\mathbf{f}$ is K **factors**, $\mathbf{u}$ is **idiosyncratic**

$$\begin{aligned} \mathbf{X} &= \mathbf{a} + \mathbf{B}\mathbf{f} + \mathbf{u}, \qquad E\mathbf{u} = 0, \quad cov(\mathbf{f}, \mathbf{u}) = 0 \quad \text{or} \\ X_j &= a_j + \mathbf{b}_j^T \mathbf{f} + u_j = a_j + b_{j1} f_1 + \cdots + b_{jK} f_K + u_j. \end{aligned}$$

★strict factor model: $\mathbf{\Sigma}_u$ diag

★approx factor model: $\mathbf{\Sigma}_u$ sparse

★multiple output regression with unknown predictor $\mathbf{f}$

Model implied covariance: $\mathbf{\Sigma} = \mathbf{B}\mathbf{\Sigma}_f\mathbf{B}^T + \mathbf{\Sigma}_u$

★low-rank + sparse

Identifiability: $\mathbf{X} = \mathbf{a} + \mathbf{B}\mathbf{f} + \mathbf{u} = \mathbf{a} + (\mathbf{B}\mathbf{H})(\mathbf{H}^{-1}\mathbf{f}) + \mathbf{u}.$

Assume ★$\text{var}(\mathbf{f}) = \mathbf{I}_K$

★columns of $\mathbf{B} = (\widetilde{\mathbf{b}}_1, \cdots, \widetilde{\mathbf{b}}_K)$ orthogonal.

High-dimensional PCA and factor analysis

Implied cov: $\mathbf{\Sigma} = \sum_{j=1}^{p} \lambda_j \xi_j \xi_j^T = \mathbf{B}\mathbf{B}^T + \mathbf{\Sigma}_u = \sum_{j=1}^{p} \lambda_j^* \xi_j^* \xi_j^{*\,T} + \mathbf{\Sigma}_u$

■ $\lambda_j^* = \|\widetilde{\mathbf{b}}_j\|^2, \quad \xi_j^* = \widetilde{\mathbf{b}}_j / \|\widetilde{\mathbf{b}}_j\|$ for $j \le K, \qquad \lambda_j = 0, \quad j > K$ are eigenvalues and eigenvectors.

Davis-Kahan Theorem $\Longrightarrow \|\xi_j - \xi_j^*\| = O(\|\mathbf{\Sigma}_u\| / \lambda_K^*)$, for $j \le K$.

Extracting factors: assume $\mathbf{a} = 0$

$$\widetilde{\mathbf{b}}_j^T \mathbf{X}_i = \|\widetilde{\mathbf{b}}_j\|^2 f_{ji} + \widetilde{\mathbf{b}}_j^T \mathbf{u}_i \approx \|\widetilde{\mathbf{b}}_j\|^2 f_{ji}, \qquad \mathbf{f_{ji}} \approx \widetilde{\mathbf{b}}_\mathbf{j}^\mathbf{T} \mathbf{X_i} / \|\widetilde{\mathbf{b}}_\mathbf{j}\|^\mathbf{2}.$$

Conclusion: factor loadings $\{\widetilde{\mathbf{b}}_j\}$ can be approximated by **PCA**: $\widetilde{\mathbf{b}}_j \approx \lambda_j^{1/2} \xi_j$ and latent factor estimated by using $f_{ji} \approx \xi_j^T \mathbf{X}_i / \lambda_j^{1/2}$.

A formal statement

<u>Pervasiveness</u>: Top K eigenvalues of $\mathbf{B}\mathbf{B}^T$ have order p, and $\|\mathbf{\Sigma}_u\|_2 = O(1)$.

★If $\{\mathbf{b}_i\}_{i=1}^p$ i.i.d., then $p^{-1}\mathbf{B}^T\mathbf{B} = p^{-1}\sum_{j=1}^p \mathbf{b}_j\mathbf{b}_j^T \to E\mathbf{b}\mathbf{b}^T$. **Hold easily**.

Prop 10.1. Under the identifiability and pervasive condition, we have

$$|\lambda_j - \|\widetilde{\mathbf{b}}_j\|_2^2| \le \|\mathbf{\Sigma}_u\|_2, \text{ for } j \le K, \qquad |\lambda_j| \le \|\mathbf{\Sigma}_u\|_2, \text{ for } j > K.$$

Furthermore,

$$\|\xi_j - \widetilde{\mathbf{b}}_j/\|\widetilde{\mathbf{b}}_j\|_2\|_2 = O(p^{-1}\|\mathbf{\Sigma}_u\|_2), \qquad \text{for all } j \le K.$$

Estimating latent factors and covariance matrix

1. Obtain an estimator $\widehat{\mu}$ and $\widehat{\mathbf{\Sigma}}$ of μ and $\mathbf{\Sigma}$
2. Do eigen-decomposition $\widehat{\mathbf{\Sigma}} = \sum_{j=1}^{p} \widehat{\lambda}_j \widehat{\xi}_j \widehat{\xi}_j^T$.
3. Estimate $\widehat{\mathbf{B}} = (\widehat{\lambda}_1^{1/2}\widehat{\xi}_1, \cdots, \widehat{\lambda}_K^{1/2}\widehat{\xi}_K)$ and $\widehat{\mathbf{f}}_i = \mathrm{diag}(\widehat{\lambda}_1, \cdots, \widehat{\lambda}_K)^{-1}\widehat{\mathbf{B}}^T(\mathbf{X}_i - \widehat{\mu})$
4. Compute $\widehat{\mathbf{\Sigma}}_u = \sum_{j=K+1}^{p} \widehat{\lambda}_j \widehat{\xi}_j \widehat{\xi}_j^T$ and its regulariztion $\widehat{\mathbf{\Sigma}}_{u,\lambda}^{\tau}$.
5. Estimate $\mathbf{\Sigma} = \mathrm{var}(\mathbf{X})$ by **POET** $\widehat{\mathbf{\Sigma}}_\lambda = \sum_{j=1}^{K} \widehat{\lambda}_j \widehat{\xi}_j \widehat{\xi}_j^T + \widehat{\mathbf{\Sigma}}_{u,\lambda}^{\tau}$.

POET for **latent** factors

★In step 4, we assume that $\mathbf{\Sigma}_u$ is sparse and applied correlation thresholding.

Prop 10.2. For sample cov with demeaned data,

$$\widehat{\mathbf{F}} = \sqrt{n} \times \text{top } K \text{ eigenvectors of } \mathbf{X}\mathbf{X}^T \quad \text{and} \quad \widehat{\mathbf{B}} = n^{-1}\mathbf{X}^T\widehat{\mathbf{F}}.$$

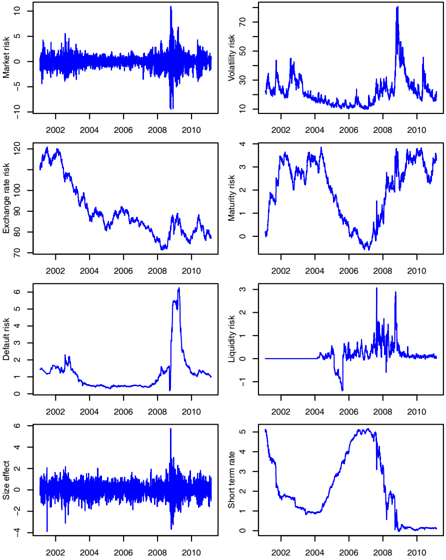
Estimating number of latent factors

Total variance: $\mathrm{tr}(\widehat{\mathbf{\Sigma}}) = \lambda_1 + \cdots + \lambda_p$.

Var explained by k^{th} PCs: $p_k = \frac{\widehat{\lambda}_k}{\widehat{\lambda}_1 + \cdots + \widehat{\lambda}_p}$; for corr $\mathbf{R}$, $p_k = \frac{\widehat{\lambda}_k}{p}$.

- ★ Scree plot: look for elbow
- ★ Adjusted Corr Threholding (ACT): $\widehat{K}_2 = \#\{j : \lambda_j(\widehat{\mathbf{R}}) > 1 + \sqrt{p/n}\}$ for corr. matrix $\widehat{\mathbf{R}}$. *(Fan, etal, 21, JASA)*
- ★ Eigenvalue ratio: $\widehat{K}_3 = \operatorname{argmax}_{j \le k_{\max}} \frac{\lambda_j(\widehat{\mathbf{\Sigma}})}{\lambda_{j+1}(\widehat{\mathbf{\Sigma}})}$ Ahn & Horenstein (13, EC), Lam & Yao (12, AOS)
- ★ Eigengap and information criteria. Bai & Ng(02)

Example: Equity Risk Factors



★market risk = returns of SP500,

★volatility risk = VIX

★exchange rate risk = value of USD

★maturity risk = yield spread between 10y and 3m treasuries,

★default risk = yield spread between AAA and BBB corporate bonds,

★liquidity risk = diff between 1m repo rates and 1m treasury bill rates,

★size effect = SP500 − R2000

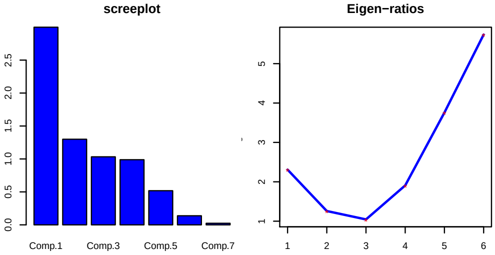
★short-term rate is LIBOR.

Market risk factors from January 29, 2001 to February 28 2011.

Combinations of known and unknown factors

Aim: Construct a risk factor, indep of SP500, based on other 7 factors.

- ★ Regress the 7 other risk factors on SP500, obtain residuals. These residual vectors are uncorrelated with SP500.
- ★ Compress these 7 residual vectors into one risk factor by first computing their correlation matrix and PCA.
- ★ Choose k by screeplot, cumul. variance explained ratios of eigenvalues.
  Ahn & Horenstein (13, EC),
  Lam & Yao (12, AOS)

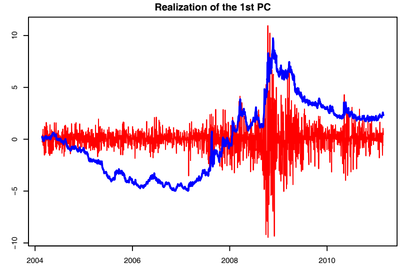


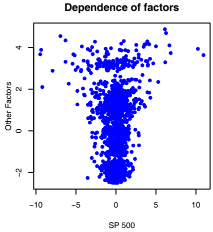
Results

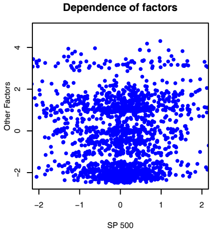
Table: Factor loadings of PCs of idiosyncratic factors to SP500

		Vol	ExR	Mat	Def	Liq	Size	Tbill
first PC	ξ_1	0.479	-0.230	0.481	0.450	0.111	0.008	-0.521
second PC	ξ_2	0.154	-0.492	-0.263	-0.049	0.779	-0.015	0.238

- 1^{st} and 2^{nd} PC explained 42.86% and 14.44% of variance. $\lambda_1 \approx 3, \lambda_2 \approx 1.01$
- Realization $\{\widehat{\xi}_1^T \mathbf{X}_t\}_{t=1}^n$ of PCA and demonstration of indep.







9.3. Covariance Estimation with Observable Factors

Least-squares estimator

1. For each $j \in [p]$, run regression $\{X_{ij} = a_j + \mathbf{b}_j^T \mathbf{f}_i + u_{ij}\}_{i=1}^n$ to obtain $\widehat{a}_j$, $\widehat{\mathbf{b}}_j$ and $\widehat{u}_{ij}$. Form $\widehat{\mathbf{B}}$ and $\{\widehat{\mathbf{u}}_i\}$.
2. Compute $\widehat{\mathbf{\Sigma}}_u$ using $\{\widehat{\mathbf{u}}_i\}$ and its correlation thresholding $\widehat{\mathbf{\Sigma}}_{u,\lambda}$.
3. Estimate $\widehat{\mathbf{\Sigma}}_\lambda = \widehat{\mathbf{B}}\widehat{\mathbf{\Sigma}}_f\widehat{\mathbf{B}}^T + \widehat{\mathbf{\Sigma}}_{u,\lambda}$.

$\lambda = 0 \Longrightarrow$ sample cov $\qquad$ $\lambda = 1 \Longrightarrow$ strick factor

$K = 0 \Longrightarrow$ regularization on sample cov.

Robust estimation of covariance matrix

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{f} \end{pmatrix}, \qquad \boldsymbol{\Sigma}_z = \begin{pmatrix} \mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^T + \boldsymbol{\Sigma}_u & \mathbf{B}\boldsymbol{\Sigma}_f \\ \boldsymbol{\Sigma}_f\mathbf{B}^T & \boldsymbol{\Sigma}_f \end{pmatrix} =: \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

1. Given a cov. input $\widehat{\boldsymbol{\Sigma}}_z$, obtain $\widehat{\boldsymbol{\Sigma}}_u = \widehat{\boldsymbol{\Sigma}}_{11} - \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21}$.
2. Apply corr thresholding to obtain $\widehat{\boldsymbol{\Sigma}}_u^{\tau}$.
3. Estimate $\boldsymbol{\Sigma}$ by $\widehat{\boldsymbol{\Sigma}}^{\tau} = \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21} + \widehat{\boldsymbol{\Sigma}}_u^{\tau}$.

■ LS estimator corresponds to $\widehat{\boldsymbol{\Sigma}}_z =$ sample covariance.

A general result *(Fan, Wang, Zhong, 19)*

Thm 10.1 Assume min and max eigenvalues of $\mathbf{\Sigma}_u \in \mathcal{C}_q(m_p)$ and $\mathbf{\Sigma}_f$ are bounded and $\|\mathbf{B}\|_{\max} \le C$, and $\lambda_K(\mathbf{\Sigma}) > cp$ for some $c > 0$.

If $\|\widehat{\mathbf{\Sigma}}_z - \mathbf{\Sigma}_z\|_\infty = O_P(\sqrt{\log p/n})$ and $Km_p\omega_n^{1-q} \to 0$ with $\omega_n = K^2\sqrt{\log p/n}$, then

$$\begin{aligned}
\|\widehat{\mathbf{\Sigma}}_u^\tau - \mathbf{\Sigma}_u\|_2 &= \|(\widehat{\mathbf{\Sigma}}_u^\tau)^{-1} - \mathbf{\Sigma}_u^{-1}\|_2 = O_P\left(m_p\omega_n^{1-q}\right), \\
\|\widehat{\mathbf{\Sigma}}^\tau - \mathbf{\Sigma}\|_{\max} &= O_P\left(\omega_n\right), \\
\|\widehat{\mathbf{\Sigma}}^\tau - \mathbf{\Sigma}\|_\Sigma &= O_P\left(\frac{K\sqrt{p}\log p}{n} + m_p\omega_n^{1-q}\right), \\
\|(\widehat{\mathbf{\Sigma}}^\tau)^{-1} - \mathbf{\Sigma}^{-1}\|_2 &= O_P\left(K^2 m_p\omega_n^{1-q}\right).
\end{aligned}$$

- relative loss $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_\Sigma = p^{-1/2}\|\mathbf{\Sigma}^{-1/2}\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1/2} - \mathbf{I}_p\|_F$ is similar to entropy loss $\mathrm{tr}(\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}) - \log|\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}| - p$.

9.4. Augmented Factor Models and Projected PCA

Fan, Liao, Wang (2016, AOS) Fan, Ke, and Liao (2019, JoE)

Examples

Example 1: We have additional clinical information that can explain partially factor loadings.

Example 2: Disaggregated macroeconomic variables should correlate with macroeconomic variables, explaining partially the latent factors.

Example 3: In equity market, we have returns of stocks, along with Fama-French factors (market, size, value, profitability, investment), momentum and firm characteristics that explaining partially the factors and loadings.

Use of Proxy Factor Information

Factor Model: $\mathbf{X}_t = \mathbf{B}\mathbf{f}_t + \mathbf{u}_t$

Augmentation: $\mathbf{f}_t = \underbrace{\mathbf{g}(\mathbf{W}_t)}_{\text{explained components}} + \underbrace{\gamma_t}_{\text{small remainders}}.$

Projection: $\widehat{\mathbf{X}} =$ fitted value $\mathbf{X}$ on $\mathbf{W}$ $\qquad \{(\mathbf{W}_t, X_{tj})\} \overset{\text{fitting}}{\Longrightarrow} \{\widehat{X}_{tj}\}$.

Fitting: ★linear regression; ★additive models; ★kernel machines

Projected PCA: Run PCA based on projected values $\widehat{\mathbf{X}}$.

Why?

$$\underbrace{E(\mathbf{X}_t|\mathbf{W}_t) = \mathbf{B}E(\mathbf{f_t}|\mathbf{W_t})}_{\mathbf{X_t^*}=\mathbf{Bf_t^*}}.$$ ⟵ **noiseless**

The estimators

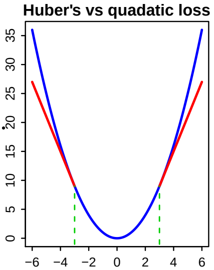
1. Compute high-dim covariance using **fitted data**.
2. Get K leading eigenvectors: $\frac{1}{\sqrt{N}}\widehat{\mathbf{B}}$
3. Estimate factors by **projecting data on PCs**: $\widehat{\mathbf{f}}_t = \frac{1}{N}\widehat{\mathbf{B}}'\mathbf{X}_t$.

★ Project fitted data on PCs: $\widehat{\mathbf{g}}(\mathbf{X}_t) = \frac{1}{N}\widehat{\mathbf{B}}'\widehat{E}(\mathbf{X}_t|\mathbf{W}_t)$.

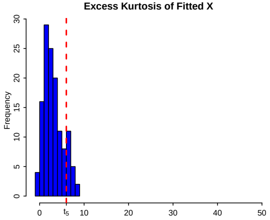
★ Compute residuals: $\widehat{\gamma}_t = \widehat{\mathbf{f}}_t - \widehat{\mathbf{g}}(\mathbf{W}_t)$.

Robust Sieve: For each asset j, fit

$$\widehat{\mathbf{b}}_j = \arg\min_{\mathbf{b}\in\mathbb{R}^J} \frac{1}{T}\sum_{t=1}^{T} \rho_{\alpha_T}\left(X_{tj} - \underbrace{\Phi(\mathbf{W}_t)}_{basis}{}'\mathbf{b}\right)$$



Fitted values: $\widehat{\mathbf{X}} = \widehat{E}(\mathbf{X}_t|\mathbf{W}_t)$



Compute: Sample covariance of $\{\widehat{\mathbf{X}}_t\}$

★Fitted variables have light tails

Forecasting Bond Risk Premia

1. **Unconditional** — **Conventional**

   ★$\mathbf{f}_t =$ extracted from $\{\mathbf{X}_t\}$ ★Regress $Y_t^{(n)}$ on $\{\mathbf{W}_t, \mathbf{f}_t\}$

2. **Conditional** — **new**

   ★$\mathbf{F}_t =$ taken from $\{\mathbf{X}_t\}$ given $\{\mathbf{W}_t\}$. ★Regress $Y_t^{(n)}$ on $\mathbf{F}_t$

Forecast model: **Linear model**

$$Y_{t+1} = \alpha + \beta' \mathbf{X}_t^* + \varepsilon_t$$

★$\mathbf{x}_t^*$: (i) $\mathbf{W}_t$; (ii)$(\mathbf{f}_t', \mathbf{W}_t')'$, (iii)$\mathbf{F}_t$; (iv)$(\mathbf{F_t}', \mathbf{W}_t')'$.

Forecast Results: Linear Model

Out-of-sample R^2

Predictors	PRP-reg				PCA-reg			
	Maturity(Year)				Maturity(Year)			
	2	3	4	5	2	3	4	5
$(\widehat{\mathbf{f}}_t', \mathbf{W}_t')'$	**37.9**	32.6	25.6	22.8	**23.9**	21.4	17.4	17.5
$\widehat{\mathbf{F}}_\mathbf{t}$	**38.1**	**32.9**	**25.7**	**23.0**	**32.6**	**28.2**	**23.3**	**19.7**
$\mathbf{W}_t$	6.1	5.5	4.7	4.5	6.1	5.5	4.7	4.5

★ Using $(\mathbf{X}_t|\mathbf{W}_t)$ to **extract** $\mathbf{F}_t$ significantly improves out-of-sample forecast, compared to directly regressing on $\{\mathbf{W}_t, \mathbf{f}_t\}$.

★ **Robust** estimations yield further improvements.

9.5. Asymptotic Properties of PCA Based Estimators

Generalization of Fan, Liu, Wang (2018, AOS)

Accuracy of Initial Estimators

Pilot estimators $\widehat{\mathbf{\Sigma}}, \widehat{\mathbf{\Lambda}}, \widehat{\mathbf{\Gamma}}$ for $\mathbf{\Sigma}$, $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_K)$ and $\mathbf{\Gamma} = (\xi_1, \ldots, \xi_K)$ satisfy

$$\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} = O_P(\sqrt{\log p/n}),$$
$$\|(\widehat{\mathbf{\Lambda}} - \mathbf{\Lambda})\mathbf{\Lambda}^{-1}\|_{\max} = O_P(\sqrt{\log p/n}),$$
$$\|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_{\max} = O_P(\sqrt{\log p/(np)}).$$

■ Conditions are modular

Regularity conditions: Top K eigenvalues are distinguished and factors are pervasive. $\|\mathbf{B}\|_{\max}, \|\mathbf{\Sigma}_u\|_2$ are bounded.

Accuracy on estimated loading matrix

Thm 10.2. (Accuracy on loading) Under accuracy + regularity condtions

$$\|\widehat{\mathbf{\Gamma}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Gamma}}^T - \mathbf{B}\mathbf{B}^T\|_{\max} = O_P(w_n),$$

$$\|\widehat{\mathbf{B}} - \mathbf{B}\|_{\max} = O_p\left(\sqrt{\frac{\log p}{n}} + \frac{1}{\sqrt{p}}\right).$$

- ★ $w_n = K^2(\sqrt{\log p/n} + \mathbf{1/\sqrt{p}})$,
- ★ $1/\sqrt{p}$ is price paid to estimate latent factors, negligible when $p \gg n \log n$
- ★ K can diverge, but results are not necessarily optimal

Properties for estimating covariance matrices

Thm10.3 (cov mat.) If $\mathbf{\Sigma}_u \in \mathcal{C}_q(m_p)$, $m_p w_n^{1-q} = o(1)$ and $\|\mathbf{\Sigma}_u^{-1}\|_2 = O(1)$,

$$\|\widehat{\mathbf{\Sigma}}_u^{\tau} - \mathbf{\Sigma}_u\|_{\max} = O_P\left(w_n\right),$$

$$\|\widehat{\mathbf{\Sigma}}_u^{\tau} - \mathbf{\Sigma}_u\|_2 = O_P\left(m_p w_n^{1-q}\right),$$

$$\|(\widehat{\mathbf{\Sigma}}_u^{\tau})^{-1} - \mathbf{\Sigma}_u^{-1}\|_2 = O_P\left(m_p w_n^{1-q}\right),$$

and for estimating $\mathbf{\Sigma}$

$$\|\widehat{\mathbf{\Sigma}}^{\tau} - \mathbf{\Sigma}\|_{\max} = O_P\left(w_n\right),$$

$$\|\widehat{\mathbf{\Sigma}}^{\tau} - \mathbf{\Sigma}\|_{\Sigma} = O_P\left(\frac{K^{3/2}p^{1/2}\log p}{n} + m_p w_n^{1-q} + K w_n/p\right),$$

$$\|(\widehat{\mathbf{\Sigma}}^{\tau})^{-1} - \mathbf{\Sigma}^{-1}\|_2 = O_P\left(K^2 m_p w_n^{1-q}\right).$$

Properties for estimating realized latent factors

Thm10.4 If $\|\widehat{\mu}-\mu\| = O_P(\sqrt{p\log p/n})$, then

$$n^{-1}\sum_{i=1}^{n}\|\widehat{\mathbf{f}}_i-\mathbf{f}_i\| = O_P\left(K/\sqrt{p}+K\sqrt{\log p/n}\right),$$

$$n^{-1}\sum_{i=1}^{n}\|\widehat{\mathbf{f}}_i-\mathbf{f}_i\|^2 = O_P\left(K^2/p+K^2\log p/n\right),$$

$$\max_{i\le n}\|\widehat{\mathbf{f}}_i-\mathbf{f}_i\| = O_P\Big(\sqrt{K/p}(\sqrt{\log p/n}+1/\sqrt{p})\max_{i\le n}\|\mathbf{X}_i-\mu\| + p^{-1}\max_{i\le n}\|\mathbf{B}^T\mathbf{u}_i\|\Big).$$

For sub-Gaussian, we can easily show

- $\max_{i\le n}\|\widehat{\mathbf{f}}_i-\mathbf{f}_i\| = O_P\left(K\sqrt{\log n/p}+K\sqrt{(\log p)(\log n)/n}\right)$.
- $\max_{i\le n, j\le p}|\widehat{\mathbf{b}}_j^T\widehat{\mathbf{f}}_i-\mathbf{b}_j^T\mathbf{f}_i| = O_P\left(K^{3/2}\sqrt{\log n/p}+K^{3/2}\sqrt{(\log p)(\log n)/n}\right)$,
- $\max_{i\le n}\|\widehat{\mathbf{u}}_i-\mathbf{u}_i\|_{\max} = O_P\left(K^{3/2}\sqrt{\log n/p}+K^{3/2}\sqrt{(\log p)(\log n)/n}\right)$.