Projected Principal Component Analysis

Yuan Liao

University of Maryland

with **Jianqing Fan and Weichen Wang**

January 3, 2015

High dimensional factor analysis and PCA

■**Factor analysis** and **PCA** are useful tools for dimension reductions.

■In high. dim. models, they are asymptotically equivalent. In particular, PCA is often used for estimating factor models.

■However, require relatively large sample size, not suitable for high-dimension-low-sample-size applications.

Conventional factor models

$$y_{it} = \lambda_i' \mathbf{f}_t + u_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T,$$

Financial asset returns:

- Over 3,000 assets are being traded in U.S. market
- Monthly data over three consecutive years contain merely 36 observations
- However, consistent estimation of loadings and factors (if unknown) requires $T \to \infty$.

Semi-parametric factor model: Connor et al. (2012):

$$y_{it} = \mathbf{g}(\mathbf{X}_i)'\mathbf{f}_t + u_{it}, \quad i = 1, ..., n, \quad t = 1, ..., T,$$

- Loadings are modeled using observed characteristics $\mathbf{X}_i$
- In financial appl., $\mathbf{X}_i$ can be firm specific var.:

market capitalization, price-earning ratio, etc

- In health studies, $\mathbf{X}_i$ can be individual char.:

weight, genetic information, etc.

- $\mathbf{g}(\cdot)$ is a nonparametric function vector.

The models assumes loading is fully explained by $\mathbf{X}_i$.

can be restrictive in many applications.

A more practical model:

$$\lambda_i = \mathbf{g}(\mathbf{X}_i) + \gamma_i$$

■γ_i: loading component that cannot be explained by $\mathbf{X}_i$, orthogonal to $\mathbf{X}_i$, $E(\gamma_i|\mathbf{X}_i) = 0$.

■Loadings depend on obser. characteristics, but not fully explained.

This paper: Projected-PCA

Model in matrix form

$$\mathbf{Y} = \{\mathbf{G}(\mathbf{X}) + \Gamma\}\mathbf{F}' + \mathbf{U}$$

Projected-PCA

- Propose **Projected-PCA** to estimate factors and loadings
- First project data $\mathbf{Y}$ on the space spanned by $\mathbf{X}$, then run PCA.
- If projection is "genuine", achieves faster convergence rate.
- Projection step "smooths away" $\mathbf{U}$ and Γ.

Remark

- Assume $\mathbf{G}(\cdot)$ additive, estimate it by sieve approximation.

Overview of the results

- Consistency is granted even if T is finite.
- If $T \to \infty$, achieve faster convergence rate if $\mathbf{G}(\cdot)$ is smooth.
- Specification tests on the loading orthogonal decomposition:

$$\Lambda = \mathbf{G}(\mathbf{X}) + \Gamma, \quad E(\Gamma|\mathbf{X}) = 0$$

  - H_0: $\mathbf{G}(\mathbf{X}) = 0$: chara. have explaining power on loadings.
  - H_0: $\Gamma = 0$: chara. fully explain loadings.
- (projection-based) Eigen-value ratio for selecting num. factors.

Projected-PCA

Main Idea

$$\mathbf{Y} = \Lambda \mathbf{F}' + \mathbf{U}, \quad E(\mathbf{U}|\mathbf{X}) = 0.$$

Normalization condition: Proj. operator $\mathcal{P}(\cdot) = E(\cdot|\mathbf{X})$.

$\mathbf{F}'\mathbf{F}/T = \mathbf{I}_K, \quad \Lambda'\mathcal{P}\Lambda$ is diagonal with district entries.

Projected-PCA:

- Operate $\mathcal{P}$ on both sides: $\mathcal{P}\mathbf{Y} = \mathcal{P}\Lambda\mathbf{F}'$,
- Projected data matrix: $\mathbf{Y}'\mathcal{P}\mathbf{Y}/T = \mathbf{F}\Lambda'\mathcal{P}\Lambda\mathbf{F}'/T$,

$$\frac{1}{T}\mathbf{Y}'\mathcal{P}\mathbf{Y}\mathbf{F} = \frac{1}{T}\mathbf{F}\Lambda'\mathcal{P}\Lambda\mathbf{F}'\mathbf{F} = \mathbf{F}\Lambda'\mathcal{P}\Lambda$$

- Columns of $\mathbf{F}$ are first K eigenvectors of $\mathbf{Y}'\mathcal{P}\mathbf{Y}/T$.

$$\frac{1}{T}\mathcal{P}\mathbf{Y}\mathbf{F} = \frac{1}{T}\mathcal{P}\Lambda\mathbf{F}'\mathbf{F} = \mathcal{P}\Lambda$$

- Both $\mathbf{F}$ and $\mathcal{P}\Lambda$ can be recovered.

The Estimator

■Observe $\mathbf{X} = \{\mathbf{X}_i\}_{i \leq p}$ i.i.d. data, each $\mathbf{X}_i$ is d-dim. r.v.

■Let $\Phi(\mathbf{X})$ be $p \times (Jd)$ matrix of basis functions, e.g., B-spline transformations of $\mathbf{X}$. J is the number of chosen basis functions.

■Let $\mathbf{P} = \Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})'$ be $p \times p$ Projection matrix.

Projected-PC estimators:

The columns of $\widehat{\mathbf{F}}/\sqrt{T}$ are the first K eigenvectors of $\mathbf{Y}'\mathbf{P}\mathbf{Y}$.

$$\widehat{\mathbf{G}}(\mathbf{X}) = \frac{1}{T}\mathbf{P}\mathbf{Y}\widehat{\mathbf{F}},$$

$$\widehat{\Lambda} = \frac{1}{T}\mathbf{Y}\widehat{\mathbf{F}}, \quad \text{Least squares on } \widehat{\mathbf{F}},$$

$$\widehat{\Gamma} = \widehat{\Lambda} - \widehat{\mathbf{G}}(\mathbf{X}) = \frac{1}{T}(\mathbf{I} - \mathbf{P})\mathbf{Y}\widehat{\mathbf{F}}.$$

Asymptotic Properties

Main Conditions

■ Identification:

$$\frac{1}{T}\mathbf{F}'\mathbf{F} = \mathbf{I}_K, \quad \Lambda'\mathbf{P}\Lambda \text{ is diagonal with distinct entries}$$

■ Genuine projection and pervasive factors:

$$c_1 < \lambda_{\min}(\Lambda'\mathbf{P}\Lambda/p) < \lambda_{\max}(\Lambda'\mathbf{P}\Lambda/p) < c_2$$

Implications:

- Projection is **genuine**: $\mathbf{X}$ has explaining power on Λ
- Factors are **pervasive**: most elements of Λ are significant.

■ Can extend to "semi-weak" factors.

Theorem In the conventional factor model $\mathbf{Y} = \Lambda \mathbf{F}' + \mathbf{U}$,

$$\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\|_F^2 = O_P(\frac{J}{p}), \quad \frac{1}{p}\|\widehat{\mathbf{G}}(\mathbf{X}) - \mathbf{P}\Lambda\|_F^2 = O_P(\frac{J}{pT} + \frac{J^2}{p^2}).$$

In the semi-parametric factor model $\Lambda = \mathbf{G}(\mathbf{X}) + \Gamma$,

$$\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\|_F^2 = O_P\left(\frac{1}{p} + \frac{1}{J^\kappa}\right),$$
$$\frac{1}{p}\|\widehat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X})\|_F^2 = O_P\left(\frac{J}{p^2} + \frac{J}{pT} + \frac{J}{J^\kappa} + \frac{J\nu_p}{p}\right).$$

- κ is the smoothing parameter, the smoother $\mathbf{G}(\cdot)$, the larger κ.
- $\nu_p = \mathrm{var}(\gamma_i)$.

Remarks

- Consistency only requires $p \to \infty$, and T can be bounded.
- When $p, T \to \infty$, faster convergence rate is achieved.
- For instance, when $\Gamma = 0$, optimal rate is

$$\frac{1}{T}\|\widehat{\mathbf{F}} - \mathbf{F}\|_F^2 = O_P\left(\frac{1}{p}\right),$$
$$\frac{1}{p}\|\widehat{\mathbf{G}}(\mathbf{X}) - \Lambda\|_F^2 = O_P\left(\frac{1}{(pT)^{1-1/\kappa}} + \frac{1}{p^{2-2/\kappa}}\right).$$

- Compared with rates of conventional factor models:

$$\frac{1}{T}\|\widetilde{\mathbf{F}} - \mathbf{F}\|_F^2 = O_P\left(\frac{1}{p} + \frac{1}{T^2}\right), \quad \frac{1}{p}\|\widetilde{\Lambda} - \Lambda\|_F^2 = O_P\left(\frac{1}{p} + \frac{1}{T}\right).$$

Loading Specification Tests

Two specification tests

$$\Lambda = \mathbf{G}(\mathbf{X}) + \Gamma, \quad E(\Gamma|\mathbf{X}) = 0$$

$H_0^1 : \mathbf{G}(\mathbf{X}) = 0$

- tests whether **X** has explaining power on loadings.
- diagnostic tool as whether or not to employ projected-PCA.

$H_0^2 : \Gamma = 0$

- tests whether **X** fully explains loadings.
- tests adequacy of semi-para. factor models in the literature.

$H_0^1: \mathbf{G}(\mathbf{X}) = 0$

■ As $\mathbf{P}\Lambda \approx \mathbf{G}(\mathbf{X})$, equivalent to testing $H_0: \mathbf{P}\Lambda = 0$.

■ rejects if S_G is large, where

$$S_G = \frac{1}{p}\mathrm{tr}(\mathbf{W}_1 \widetilde{\Lambda}' \mathbf{P} \widetilde{\Lambda}), \quad \mathbf{W}_1 = (\frac{1}{p}\widetilde{\Lambda}'\widetilde{\Lambda})^{-1}.$$

■ $\widetilde{\Lambda}$ is traditional PC estimator, as the projected-PCA is inconsistent under H_0^1.

■ Under H_0^1, as $p, T, J \to \infty$,

$$\frac{pS_G - JdK}{\sqrt{2JdK}} \to^d N(0,1),$$

$H_0^2 : \Gamma = 0$

- As $\Gamma \approx \Lambda - \mathbf{P}\Lambda$, equivalent to testing $H_0 : \mathbf{P}\Lambda = \Lambda$.
- rejects if S_Γ is large, where

$$S_\Gamma = \mathrm{tr}(\widehat{\Lambda}(\mathbf{I}-\mathbf{P})\Sigma_u^{-1}(\mathbf{I}-\mathbf{P})\widehat{\Lambda})$$

- Assume Σ_u to be diagonal. Replace with a diagonal estimator.
- Can extend to non-diagonal but sparse Σ_u.
- Under H_0^2, as $p, T, J \to \infty$,

$$\frac{TS_\Gamma - pK}{\sqrt{2pK}} \to^d N(0,1).$$

Testing on real data

■Data: daily ex. returns of S&P 500 constituents 2005-2013.

■Four characteristics $\mathbf{X}$ as in Connor et al. 12:

size, value, momentum and volatility.

■Tests are based on rolling windows.

Results:

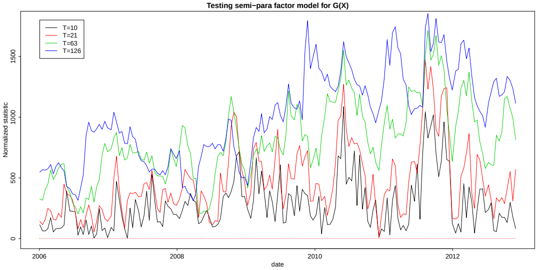
■Strong evidence of explaining powers of $\mathbf{X}$ on loadings;

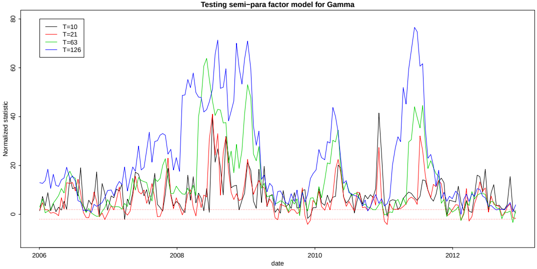
providing theoretical basis for projected-PCA.

■For a majority of period, $\Gamma = 0$ is rejected; assets'

characteristics cannot fully explain markets' betas.

Figure: Normalized statistics. Upper: S_G; Lower: S_Γ





Estimating the Number of Factors

■Strengths of factors are affected by eigenvalues of $\mathbf{G}(\mathbf{X})\mathbf{G}(\mathbf{X})'$. $\mathbf{G}(\mathbf{X})\mathbf{G}(\mathbf{X})'$ has K very spiked eigenvalues.

■$\mathcal{P}\mathbf{Y} = \mathbf{G}(\mathbf{X})\mathbf{F}'$ implies

$$\frac{1}{T}\mathcal{P}\mathbf{Y}(\mathcal{P}\mathbf{Y})' = \mathbf{G}(\mathbf{X})\mathbf{G}(\mathbf{X})'$$

Eigenvalues of $\mathbf{G}(\mathbf{X})\mathbf{G}(\mathbf{X})'$ are the same as those of $\mathbf{Y}\mathcal{P}\mathbf{Y}'/T$.

■Hence, employ **eigenvalue-ratio method** (Ahn and Horenstein 13 and Lam and Yao 12) on the **projected data** $\mathbf{PY}$:

$$\widehat{K} = \arg \max_{1 \le k \le [\min\{p,T\}/2]} \frac{\lambda_k(\mathbf{Y}'\mathbf{PY})}{\lambda_{k+1}(\mathbf{Y}'\mathbf{PY})}.$$

■Allow semi-weak factors:

$$c < \lambda_{\min}(p^{-\alpha}\mathbf{G}(\mathbf{X})'\mathbf{G}(\mathbf{X})) < \lambda_{\max}(p^{-\alpha}\mathbf{G}(\mathbf{X})'\mathbf{G}(\mathbf{X})) < C$$

for some $\alpha \in (0, 1]$.

■Conventionally, $\alpha = 1$, "strong factors".

Theorem Under regularity conditions, as $p, T, J \to \infty$,

$$P(\widehat{K} = K) \to 1.$$

Numerical Studies

Design 1

- Data of 337 stocks from S&P 500 are collected.
- Use four char. and three factors.
- “True values” of GDP are calibrated from the data.

Figure: P-PCA of $\|\widehat{\mathbf{G}} - \mathbf{G}\|$, P-PCA of $\|\widehat{\Gamma} - \Gamma\|$, and regular PCA of $\|\widetilde{\Lambda} - \Lambda\|$

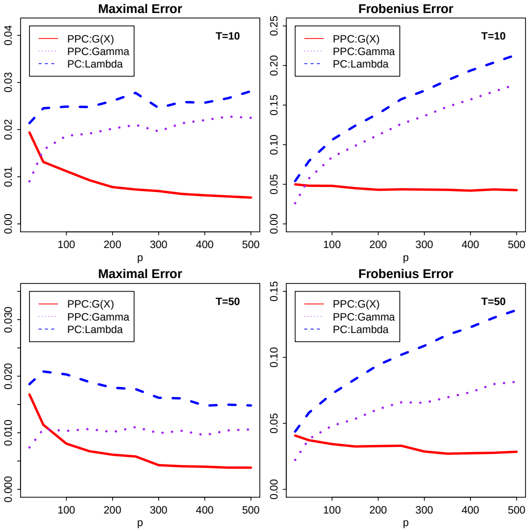
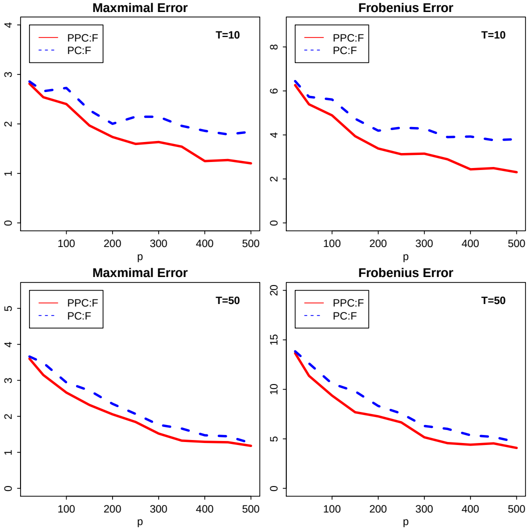


Figure: P-PCA of $\|\widehat{\mathbf{F}} - \mathbf{F}\|$, and regular PCA of $\|\widetilde{\mathbf{F}} - \mathbf{F}\|$



Design 2

- Simulate one characteristic and three factors.

$$\mathbf{g} = (x, x^2 - 1, x^3 - 2x), \quad \Gamma = 0.$$

- Compare three methods for estimating loadings:

Projected-PCA, PCA, and least squares w/ known factors (SLS).

- Compare two methods for estimating K:

on projected data and on non-projected data.

Results: Projected-PCA performs:

- significantly better than regular PCA.
- as well as if the factors are known when p is large.
- more accurately in estimating K.

Figure: $\|\widehat{\mathbf{G}}(\mathbf{X}) - \mathbf{G}(\mathbf{X})\|$ of Projected-PCA, PCA, and SLS

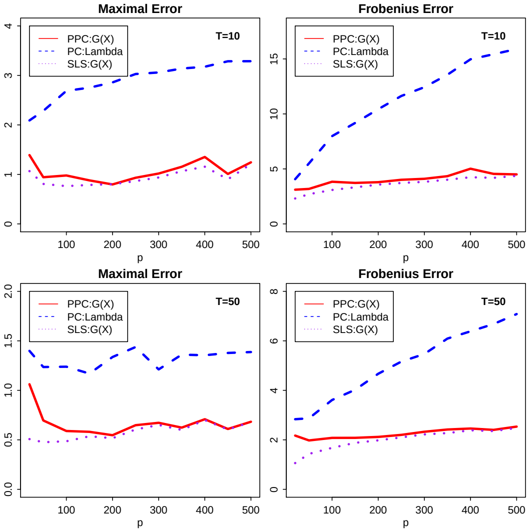
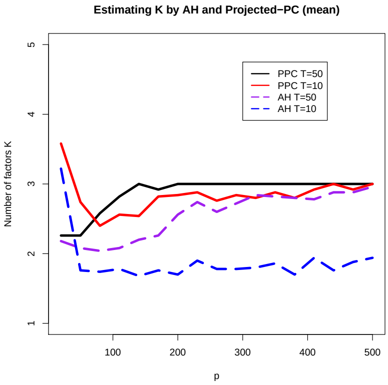
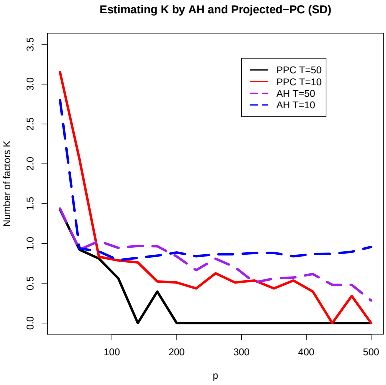


Figure: Estimating $K = 3$ using Projected-PCA with Ahn and Horenstein 13 (non-projected data), 50 replications

mean

standard deviation





Summary

Semi-parametric Factor model

- Loadings depend on observed characteristics.
- More flexible than existing model specifications.
- Verified by empirical studies.

Projected-PCA:

- Apply PCA on projected data.
- Consistency is granted even under finite sample size.
- Faster rate of convergence

Loading Specification Tests:

- Tests whether char. have explaining powers on loadings
- Tests whether char. fully explain loadings.