Characteristics based Factor Models

Yuan Liao

Rutgers University

based on joint works with

Jianqing Fan, Yuan Ke, Weichen Wang

December 17, 2016

- Projected principal components analysis in factor models (with Fan and Wang), *Ann. Statist.* 2016
- Robust factor models with covariates (with Fan and Ke), *manuscript.* 2016+

Factor models

$$y_{it} = \lambda_i' f_t + u_{it}, \quad i \leq N, t \leq T$$

- When the dim is large, often estimated by PCA.
- In high-dim. spiked models, they are asympt. equivalent.
- Large literature in economics, finance, statistics,

  Chamberlain and Rothschild (83), Stock and Watson (02) , Bai and Ng (02), Forni, Hallin, Lippi, Reichlin (00), ...

Use of Covariate information

Covariate-based factor models:

- ★ Depend on covariates through either λ_i, or $\mathbf{f}_t$ or both.
- ★ **Through loadings**:

$$\lambda_i = \underbrace{\mathbf{g}(\mathbf{X}_i)}_{\text{explained components}} + \underbrace{\gamma_i}_{\text{remainders}},$$

Estimation is improved when T is small.

- ★ **Through factors**:

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

Estimation is improved when N is small.

Key Assumptions about X

★$\mathbf{X}$ is **strongly associated with** $\mathbf{Y}$:

- Let $\widehat{\mathbf{Y}}$ be the fitted value when regressing $\mathbf{Y}$ on $\mathbf{X}$, then

$$\text{The first } K \text{ eigenvalues of } (\frac{1}{NT}\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}') \text{ is bounded away from } 0, \infty.$$

- This means:

Through loadings:

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{g}(\mathbf{X}_i)\mathbf{g}(\mathbf{X}_i)' \text{ does not degenerate}$$

Through factors:

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(\mathbf{X}_t)\mathbf{g}(\mathbf{X}_t)' \text{ does not degenerate}$$

★$\mathbf{X}$ is mean-independent of $(\mathbf{u}_t, \gamma)$.

Through loadings

Semi-parametric factor model: (*Connor & Linton, 07; +Mathias 12*):

$$y_{it} = \mathbf{g}(\mathbf{X}_i)'\mathbf{f}_t + u_{it}, \quad \text{or} \quad \lambda_i = g(\mathbf{X}_i)$$

- Loadings are modeled using covariates $\mathbf{X}_i$.
  - **Finance**: $\mathbf{X}_i =$ firm-specific variables: size, PE, PS ratios, etc
  - **Health**: $\mathbf{X}_i =$ can be individual char. weight, genetic information, etc.
- Loadings are fully explained by $\mathbf{X}_i$, can be restrictive.

Flexible semiparametric model

Fan, Liao and Wang (16, AOS):

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

- Loadings are explained partially.
- Two-step estimation:

  Step 1 Project $\mathbf{Y}$ onto the space of $\mathbf{X}$, obtain $\widehat{\mathbf{Y}}$.

  Step 2 Run PCA on the fitted data $\widehat{\mathbf{Y}}$.
- Results:
  - Factors are consistently estimated when T is finite.
  - Faster rate of convergences when T grows.
  - Specification tests: $\mathbf{g} = 0$ and $\gamma = 0$.

The Idea

- Projection matrix: $\mathbf{P} = \Phi(\mathbf{X})(\Phi(\mathbf{X})'\Phi(\mathbf{X}))^{-1}\Phi(\mathbf{X})'$.
- $$\mathbf{P}\mathbf{Y}_t \approx (\mathbf{P}\mathbf{G} + \underbrace{\mathbf{P}\gamma}_{\approx 0})^T \mathbf{f}_t + \underbrace{\mathbf{P}\mathbf{u}_t}_{\approx 0} \approx (\underbrace{\mathbf{P}\mathbf{G}}_{\approx \mathbf{G}})^{\mathbf{T}} \mathbf{f}_{\mathbf{t}}.$$

So

$$\frac{1}{T}\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}' \approx \mathbf{G}\mathrm{cov}(\mathbf{f}_t)\mathbf{G}'$$

- γ and $\mathbf{u}_t$ are "projected off".

$$\frac{1}{T}\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}'\mathbf{G} \approx \mathbf{G}\mathrm{cov}(\mathbf{f}_t)\mathbf{G}'\mathbf{G}$$

so columns of **G** are approx. space of the leading eigenvectors

Through factors

Covariates may also appear in factors.

$$\mathbf{f}_t = \mathbf{g}(\mathbf{X}_t) + \gamma_t.$$

- Factors may be partially explained by covariates $\mathbf{X}_t$.
  - **Finance**: $\mathbf{X}_t =$ Fama-French factors, etc
  - **Macro-forecasts**: $\mathbf{X}_t =$ consumption-wealth variable, financial ratios, and term spread, etc.
- Fama-French factors cannot fully explain $\mathbf{f}_t$ (Fama-French 15).

Identification

$$E(\mathbf{Y}_t|\mathbf{X}_t) = \Lambda\mathbf{g}(\mathbf{X}_t), \quad \mathbf{g}(\mathbf{X}_t) = E(\mathbf{f}_t|\mathbf{X}_t).$$

Hence

$$\Sigma := E\{E(\mathbf{Y}_t|\mathbf{X}_t)E(\mathbf{Y}_t|\mathbf{X}_t)'\} = \Lambda \underbrace{E\{\mathbf{g}(\mathbf{X}_t)\mathbf{g}(\mathbf{X}_t)'\}}_{\Sigma_g} \Lambda'$$

Normalization condition: $\frac{1}{N}\Lambda'\Lambda = \mathbf{I}$, and Σ_g is diagonal.

Then

$$\frac{1}{N}\Sigma\Lambda = \Lambda\Sigma_g, \quad \mathbf{g}(\mathbf{X}_t) = \frac{1}{N}\Lambda' E(\mathbf{Y}_t|\mathbf{X}_t)$$

★ Columns of Λ are the leading eigenvectors of Σ.

★ Exact identification, as opposed to the "asym. ident." in the literature.

★ Identified up to rotations without normalization conditions.

The estimators

★ Estimate high-dim cov $\Sigma := E\{E(\mathbf{Y}_t|\mathbf{X}_t)E(\mathbf{Y}_t|\mathbf{X}_t)'\}$. e.g.,

$$\widehat{\Sigma} = \frac{1}{NT}\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}}$$

★ The columns of $\frac{1}{\sqrt{N}}\widehat{\Lambda}$ are the eigenvectors corresponding to the first K eigenvalues of $\widehat{\Sigma}$. (usual methods use $\frac{1}{NT}\mathbf{Y}'\mathbf{Y}$)

★ $\widehat{\mathbf{g}}(\mathbf{X}_t) := \frac{1}{N}\widehat{\Lambda}'\widehat{E}(\mathbf{Y}_t|\mathbf{X}_t)$.

★ $\widehat{\mathbf{f}}_t = \frac{1}{N}\widehat{\Lambda}'\mathbf{Y}_t$.

★ $\widehat{\gamma}_t = \widehat{\mathbf{f}}_t - \widehat{\mathbf{g}}(\mathbf{X}_t)$.

General Consistency

Theorem 1: Suppose under the spectral norm,

$$\|\widehat{\Sigma} - \Sigma\| = o_P(N),$$

then there is $\mathbf{H}$, so that

$$\frac{1}{N}\|\widehat{\Lambda} - \Lambda\mathbf{H}\|_F^2 = o_P(1).$$

In addition, if the normalization condition holds,

$$\mathbf{H} = \mathbf{I}.$$

Heavy-tailed Panel Data

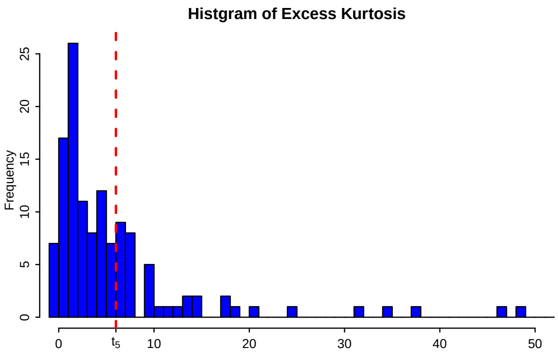
$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\widehat{E}(\mathbf{Y}_t|\mathbf{X}_t)\widehat{E}(\mathbf{Y}_t|\mathbf{X}_t)', \quad \text{where } \widehat{E}(\mathbf{Y}_t|\mathbf{X}_t) = \widehat{\mathbf{B}}\Phi(\mathbf{X}_t).$$

– $\widehat{\Sigma} = \frac{1}{NT}\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}}$: use $\widehat{\mathbf{B}}$ as the least squares.

– Can use a better estimator for $\widehat{\mathbf{B}}$ to protect against heavy tails.

– Many series in the panel data are heavy tailed.

Example: Monthly data of 131 macroeconomic series 1964-2003 (e.g. Ludvigson and Ng 2009, 2010).



★43 series have tails heavier than t_5

Robust estimation of Σ

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \widehat{E}(\mathbf{Y}_t|\mathbf{X}_t)\widehat{E}(\mathbf{Y}_t|\mathbf{X}_t)', \quad \text{where } \widehat{E}(\mathbf{Y}_t|\mathbf{X}_t) = \widehat{\mathbf{B}}\Phi(\mathbf{X}_t).$$

★For some deterministic sequence $\alpha_T \to \infty$,

$$\widehat{\mathbf{b}}_i = \arg\min_{\mathbf{b}\in\mathbb{R}^J} \frac{1}{T}\sum_{t=1}^{T} \rho\left(\frac{y_{it} - \Phi(\mathbf{X}_t)'\mathbf{b}}{\alpha_T}\right), \quad \widehat{\mathbf{B}} = (\widehat{\mathbf{b}}_1, ..., \widehat{\mathbf{b}}_N)'.$$

Huber loss(Huber, 1964):

$$\rho(z) = \begin{cases} z^2, & |z| < 1 \\ 2|z| - 1, & |z| \geq 1. \end{cases}$$

Study of US Bond Risk Premia

Econometric Motivation

Can US bond risk premia be explained by large macroeconomic panel data?

Approaches based on diffusion index models

- Regress on $\{\mathbf{X}_t, \mathbf{f}_t\}$ (Stock, Watson 02; Ludvigson, Ng 09, 10)

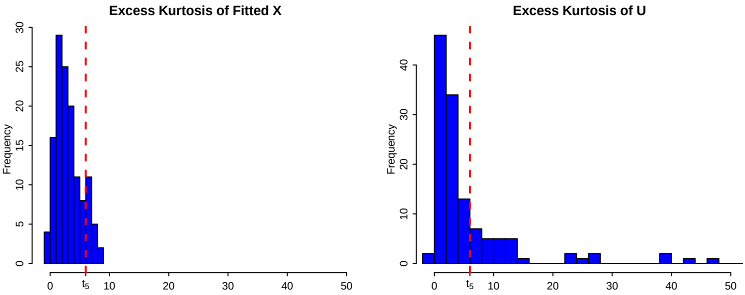
Our Discoveries

- Using $\mathbf{X}_t$ to **explain** $\mathbf{f}_t$ (instead of as predictors) significantly improves out-of-sample forecast, compared to directly regressing on $\{\mathbf{X}_t, \mathbf{f}_t\}$.
- Robust estimations yield further improvements.

Data and Model

- ★ $y_t^{(n)}$: US bond risk premia with maturity of n years, $n = \{2, \cdots, 5\}$.
- ★ $\mathbf{Y}_t$: Monthly data of 131 macroeconomic series 1964-2003 (e.g. Ludvigson and Ng 2009, 2010).
- ★ $\mathbf{X}_t$: 8 observables used to describe the co-movement of the macroeconomic activities (e.g. NBER, 08; Stock, Watson 10).

Robustly Fitted Data



★The fitted data $\widehat{E}(\mathbf{Y}_t|\mathbf{X}_t)$ are no longer severely heavy-tailed.

★The estimated idiosyncratic errors preserve the heavy-tailed behavior.

Raw data: check here

Forecast models

Linear model

$$y_{t+1} = \alpha + \beta' \mathbf{z}_t + \varepsilon_t$$

Multi-index model (Li, 1991; Fan, Xue, Yao, 2015)

$$y_{t+1} = \alpha + h(\psi_1' \mathbf{z}_t, \cdots, \psi_L' \mathbf{z}_t) + \varepsilon_t$$

- $\mathbf{z}_t$: (i) $\mathbf{X}_t$; (ii)$\mathbf{f}_t$; (iii)$(\mathbf{f}_t', \mathbf{X}_t')'$.
- $L < dim(\mathbf{X}_t)$!
- $\mathbf{f}_t$ is obtained from 131 macro time series.

Rolling Window Forecast

- Forecast y_{T+t+1} using the data of the previous $T = 240$ months.
- The forecast performance is assessed by the out-of-sample R^2

$$R^2 = 1 - \frac{\sum_{t=0}^{239} (y_{T+t+1} - \hat{y}_{T+t+1|T+t})^2}{\sum_{t=0}^{239} (y_{T+t+1} - \bar{y}_t)^2},$$

where $\bar{y}_t$ is the sample mean of y_t over the sample period $[1 + t, T + t]$.

Forecast Results: Linear Model

★ $y_{t+1} = \alpha + \beta' \mathbf{z}_t + \varepsilon_t$

Table: out-of-sample R^2, the larger the better

$\mathbf{z}_t$	proposed method				PCA			
	Maturity(Year)				Maturity(Year)			
	2	3	4	5	2	3	4	5
$(\widehat{\mathbf{f}}_t', \mathbf{X}_t')'$	37.9	32.6	25.6	22.8	23.9	21.4	17.4	17.5
$\widehat{\mathbf{f}}_t$	38.1	32.9	25.7	23.0	32.6	28.2	23.3	19.7
$\mathbf{X}_t$	6.1	5.5	4.7	4.5	6.1	5.5	4.7	4.5

Forecast Results: Multi-index Model

$$y_{t+1} = \alpha + h(\psi_1' \mathbf{z}_t, \cdots, \psi_L' \mathbf{z}_t) + \varepsilon_t$$

$\mathbf{z}_t$	proposed method				PCA			
	Maturity(Year)				Maturity(Year)			
	2	3	4	5	2	3	4	5
$(\mathbf{f}_t', \mathbf{X}_t')'$	41.7	39.0	35.6	34.1	30.8	26.3	24.6	22.0
$\mathbf{f}_t'$	41.2	39.1	35.2	34.1	34.5	32.1	27.3	23.7
$\mathbf{X}_t$	13.6	10.8	10.0	6.8	13.6	10.8	10.0	6.8

★More results in our paper.

Findings

1. $\mathbf{X}_t$ contains strong explanatory powers of the latent factors.
2. The gain is more substantial when incorporate $\mathbf{X}_t$ to estimate $\mathbf{f}_t$ than only use it for forecasting.
3. Robust estimations yield significant impovements.
4. The multi-index models out-performs linear models.

Conclusion

Projected-PCA:

- Semi-parametric factor model
- Apply PCA on projected data.

Robust Proxy-regressed Method

- Apply PCA on **Robustly** fitted data.
- Little price under light-tails and Big gain under heavy-tails.
- Faster rate of convergence.