State-Varying Factor Models of Large Dimensions

Markus Pelger and Ruoxuan Xiong

Stanford University

European Econometric Society Meeting
August 27, 2018
Motivation

- Conventional large-dimensional latent factor model assumes the exposures to factors (factor loadings) are constant over time.
- Observation: Asset prices’ exposures to the market (and other risk factors) are time-varying.
- Example: Term-structure factor exposure is different in recessions and booms.

**Figure**: PCA Factor Loadings for Treasuries in Boom and Recession.

(a) Level Factor  (b) Slope Factor  (c) Curvature Factor
This paper

Research Question:

1. Find latent factors and loadings that are state-dependent.
2. Test if factor model is state-dependent.

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 cross-section units with many time observations
3. Factor structure can be time-varying as a general non-linear function of the state process
Contribution of this paper

**Contribution**

- **Theoretical**
  - PCA estimator combined with kernel projection for factors, state-varying factor loadings and common components
  - Asymptotic inferential theory for estimators for $N, T \to \infty$:
    - consistency
    - normal distribution and standard errors
  - Test for state-dependency of latent factor model
    - Generalized correlation test statistic detects for which states model changes
    - Non-standard superconsistency

- **Empirical**
  - State-dependency of factor loadings in US Treasury securities
Literature (partial list)

- Large-dimensional factor models with constant loadings
  - Bai (2003): Distribution theory
  - Fan et al. (2013): Sparse matrices in factor modeling

- Large-dimensional factor models with time-varying loadings
  - Su and Wang (2017): Local time-window
  - Pelger (2018), Aït-Sahalia and Xiu (2017): High-frequency
  - Fan et al. (2016): Projected PCA

- Large-dimensional factor models with structural breaks
  - Stock and Watson (2009): Inconsistency
The Model

State-varying factor model

- $X_{it}$ is the observed data for the $i$-th cross-section unit at time $t$
- State variable $S_t$ at time $t$

$$X_{it} = \Lambda_i(S_t) F_t + e_{it} \quad i = 1, \ldots N, \ t = 1, \ldots T$$

- $N$ cross-section units (large), time horizon $T$ (large)
- $r$ systematic factors (fixed)
- Factors $F$, loadings $\Lambda(S_t)$, idiosyncratic components $e$ are unknown
- Data $X$ and state process $S_t$ observed
The Model

Examples (with one factor) equivalent to multi-factor representation

- Loadings linear in state: \( \Lambda_i(S_t) = \Lambda_{i,1} + \Lambda_{i,2}S_t \)
  \[ X_{it} = \Lambda_{i,1}F_t + \Lambda_{i,2}(S_tF_t) + e_{it} \]
  
- Loadings nonlinear in discrete state: \( \Lambda_i(S_t) = g_i(S_t), \) \( S_t \in \{s_1, s_2\} \)
  \[ X_{it} = g_i(s_1)\mathbb{1}_{\{S_t=s_1\}}F_t + g_i(s_2)\mathbb{1}_{\{S_t=s_2\}}F_t + e_{it} \]

Our model

- Loadings nonlinear in non-discrete state: \( \Lambda_i(S_t) = g_i(S_t) \) with continuous distribution function for \( S_t \)
  \( \Rightarrow \) Cumbersome/No multi-factor representation
The Model: Main Assumptions

Approximate state-varying factor model

- Systematic factors explain a large portion of the variance
- Idiosyncratic risk is nonsystematic: Weak time-series and cross-sectional correlation
- State: recurrent (infinite observations around the state to condition on) with continuous stationary PDF
- Factor Loadings: deterministic functions of the state and the functions are Lipschitz continuous (observations in the nearby state are useful)

$$\exists C, \| \Lambda_i(s + \Delta s) - \Lambda_i(s) \| \leq C |\Delta s|$$
The observed state process is a noisy approximation of the underlying state process (e.g. omitted state)

\[ X_{it} = (\Lambda_i(S_t) + \varepsilon_{it})^T F_t + e_{it} \quad i = 1, 2, \ldots, N \text{ and } t = 1, 2, \ldots, T \]

or in vector notation

\[
\begin{align*}
\begin{pmatrix} X_t \end{pmatrix}_{N \times 1} &= \begin{pmatrix} \Lambda(S_t) \end{pmatrix}_{N \times r} \begin{pmatrix} F_t \end{pmatrix}_{r \times 1} + \begin{pmatrix} \varepsilon_t \end{pmatrix}_{N \times r} \begin{pmatrix} F_t \end{pmatrix}_{r \times 1} + \begin{pmatrix} e_t \end{pmatrix}_{N \times 1} \\
&= \Lambda(S_t)F_t + \psi_t + e_t
\end{align*}
\]

- Under weak conditions noise in state process can be treated like idiosyncratic noise.

⇒ All results hold!
The Model: Intuition

Intuition for Estimation

- **Constant loadings:**
  Loadings are principal components of covariance matrix

\[ \text{Cov}(X_t) = \Lambda \text{Cov}(F_t) \Lambda^\top + \text{Cov}(e_t). \]

- **State-varying loadings:**
  Loadings for \( S_t = s \) are principal components of covariance matrix conditioned on the state \( S_t = s \):

\[ \text{Cov}(X_t|S_t = s) = \Lambda(s) \text{Cov}(F_t|S_t = s) \Lambda(s)^\top + \text{Cov}(e_t|S_t = s). \]

⇒ **Intuition:** Estimate conditional covariance matrix \( \text{Cov}(X_t|S_t = s) \) with kernel projection and apply PCA to it.
Objective function and nonparametric estimation

The estimators minimize mean squared error conditioned on state:

$$\hat{F}^s, \hat{\Lambda}(s) = \arg\min_{F^s, \Lambda(s)} \frac{1}{N T(s)} \sum_{i=1}^{N} \sum_{t=1}^{T} K_s(S_t)(X_{it} - \Lambda_i(s)'F_t)^2$$

- Kernel function $K_s(S_t) = \frac{1}{h} K\left(\frac{S_t - s}{h}\right)$ (e.g. $K(u) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{u^2}{2}\}$)
- $T(s) = \sum_{t=1}^{T} K_s(S_t), \frac{T(s)}{T} \xrightarrow{d} \pi(s)$ (stationary density of $S_t = s$)
- Bandwidth parameter $h$ determines local “state window”
The Model: Nonparametric Estimation

Nonparametric estimation

- Project square root of kernel on the data and factors
  \[ X_{it}^s = K_1^{1/2}(S_t)X_{it} \quad F_t^s = K_1^{1/2}(S_t)F_t \]

- PCA solves optimization problem
  \[ \hat{F}^s, \hat{\Lambda}(s) = \arg \min_{F^s, \Lambda(s)} \frac{1}{NT(s)} \sum_{i=1}^N \sum_{t=1}^T (X_{it}^s - \Lambda_i(s)'F_t^s)^2 \]

⇒ Apply PCA to conditional covariance matrix

- \( \hat{F}^s \) are the eigenvectors corresponding to top k eigenvalues of estimated conditional covariance matrix \( \frac{1}{NT(s)}(X^s)'X^s \)

- \( \hat{\Lambda}(s) \) are coefficients from regressing \( X^s \) on \( \hat{F}^s \)
The Model: Nonparametric Estimation

Major challenge: Bias term from using nearby states

\[
X_t^s = \Lambda(S_t)F_t^s + e_t^s = \underbrace{\Lambda(s)F_t^s + e_t^s}_{\bar{X}_t^s} + \underbrace{(\Lambda(S_t) - \Lambda(s))F_t^s}_{\Delta X_t^s}.
\]

- \(\Delta X_{it}^s = \Lambda_i(S_t)F_t^s - \Lambda_i(s)F_t^s = O_p(h)\)
- Kernel bias complicates problem and lowers convergence rates

Theorem: Consistency

Assume \(N, Th \to \infty\) and \(\delta_{NT,h} h \to 0\) with \(\delta_{NT,h} = \min(\sqrt{N}, \sqrt{Th})\):

\[
\delta_{NT,h}^2 \left( \frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_t^s - (H^s)^T F_t^s \right\|^2 \right) = O_p(1)
\]

\[
\delta_{NT,h}^2 \left( \frac{1}{N} \sum_{i=1}^N \left\| \hat{\Lambda}_i(s) - (H^s)^{-1} \Lambda_i(s) \right\|^2 \right) = O_p(1)
\]

for known full rank matrix \(H^s\)
Limiting Distribution of Estimated Factors

Theorem (Factors)

Assume $\sqrt{Nh}/(Th) \to 0$, $Nh \to \infty$ and $Nh^2 \to 0$. Then

$$
\sqrt{N} \left( K_s^{-1/2}(S_t) \hat{F}_t^s - (H^s)'F_t \right)
= (V_s^r)^{-1} \frac{(\hat{F}_t^s)'F_t}{T} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_i(s)e_{it} + o_p(1)
\xrightarrow{D} N(0, (V^s)^{-1} Q_t^s \Gamma_t^s (Q^s)' (V^s)^{-1})
$$

- Rotation matrix $H^s = \frac{\Lambda(s)'\Lambda(s)}{N} \frac{(F_t^s)'\hat{F}_t^s}{T} (V_s^r)^{-1}$
- $K_s^{-1/2}(S_t) \hat{F}_t^s$ converges to some rotation of $F_t$ at rate $\sqrt{N}$
- Efficiency mainly depends on asymptotic distribution of $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_i(s)e_{it}$
Asymptotic Results

Limiting Distribution of Estimated Factor Loadings

**Theorem (Loadings)**

Assume $\sqrt{Th}/N \to 0$, $Th \to \infty$, and $Th^3 \to 0$. Then

$$
\sqrt{Th}(\hat{\Lambda}_i(s) - (H^s)^{-1}\Lambda_i(s)) = (V_r^s)^{-1}(\hat{F}^s)' F^s \frac{\Lambda(s)'^{\prime} \Lambda(s)}{Th} \frac{\sqrt{Th}}{N} \sum_{t=1}^{T} F_t^s e_t^s + o_p(1)
$$

$$
D \Rightarrow N(0, ((Q^s)')^{-1}\Phi_i^s (Q^s)^{-1})
$$

- $\hat{\Lambda}_i(s)$ converges to some rotation of $\Lambda_i(s)$ at rate $\sqrt{Th}$.
- Efficiency mainly depends on asymptotic distribution of $
\frac{\sqrt{Th}}{T(s)} \sum_{t=1}^{T} F_t^s e_t^s = \frac{\sqrt{Th}}{T(s)} \sum_{t=1}^{T} K_s(S_t) F_t e_{it}
$.
Asymptotic Results

Limiting Distribution of Common Component

Theorem (Common Components)

Assume $Nh \to \infty$, $Th \to \infty$, $Nh^2 \to 0$ and $Th^3 \to 0$. Then for each $i$

$$
\delta_{NT,h}(\hat{C}_{it,s} - C_{it,s}) = \frac{\delta_{NT,h}}{\sqrt{N}} \Lambda_i(s) \Sigma^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Lambda_i(s)e_{it} \right) + \frac{\delta_{NT,h}}{\sqrt{Th}} F_t' \Sigma^{-1} F_{ts} \left( \frac{\sqrt{Th}}{T(s)} \sum_{t=1}^{T} F^s_t e^s_{it} \right) + o_p(1)
$$

- $\delta_{NT,h} = \min(\sqrt{N}, \sqrt{Th})$
- Define $C_{it,s} = F_t' \Lambda_i(s)$ and $\hat{C}_{it,s} = (\frac{\hat{F}^s_t}{K^{1/2}_s(S_t)})' \hat{\Lambda}_i(s)$
- If $N/(Th) \to 0$, $\Lambda_i(s)e_{it}$ dominates
- If $Th/N \to 0$, $F^s(t)e^s_{it}$ dominates
Test for constancy: Generalized correlation test

Consider loadings in two states $\Lambda_1 = \Lambda(s_1)$ and $\Lambda_2 = \Lambda(s_2)$. Test for

$\mathcal{H}_0 : \Lambda_1 = \Lambda_2 G$ for some full rank square matrix $G$

$\mathcal{H}_1 : \Lambda_1 \neq \Lambda_2 G$ for any full rank square matrix $G$

- Generalized correlation, defined as $\rho$ invariant of $G$

$$
\rho = \text{trace} \left\{ \left( \frac{\Lambda_1^T \Lambda_1}{N} \right)^{-1} \left( \frac{\Lambda_1^T \Lambda_2}{N} \right) \left( \frac{\Lambda_2^T \Lambda_2}{N} \right)^{-1} \left( \frac{\Lambda_2^T \Lambda_1}{N} \right) \right\}
$$

- $\hat{\rho}$ estimated $\rho$ and $r$ is #factors

- Equivalent to test $\mathcal{H}_0 : \rho = r$ and $\mathcal{H}_1 : \rho < r$
Generalized Correlation

Theorem: Generalized correlation test

Assume $\sqrt{N/(Th)} \to 0$, $Nh \to \infty$, $Th \to \infty$, $\sqrt{Th/N} \to 0$, $Nh^2 \to 0$ and $NTh^3 \to 0$:

$$\sqrt{NTh}(\hat{\rho} - r - \hat{\xi}^\top \hat{b}) \xrightarrow{d} N(0, \Omega)$$

- $\hat{\xi}^\top \hat{b}$ bias term with feasible estimates $\hat{b}$ and $\hat{\xi}$
- feasible estimator for asymptotic covariance $\hat{\Omega}$
- Superconsistent rate $\sqrt{NTh}$ (corner case)
- $h \in [1/T^{1/2}, 1/T^{3/4}]$: combinations of $N$ and $T$ exist to satisfy the rate conditions
Empirical Applications

- US Treasury Securities Yields from 2001-07-31 to 2016-12-01: \( N = 11, \ T = 2832 \): 1, 3, 6 mo., 1, 2, 3, 5, 7, 10, 20, 30 yr.
- State: Log-normalized VIX
- Generalized correlation: \( \hat{\rho}(\Lambda(Boom), \Lambda(Recession)) = 2.6352 \)
  \( \Rightarrow \) reject \( \rho \approx 3 \) for \( \Lambda(Boom) \approx \Lambda(Recession) \)

(a) Log-normalized VIX

(b) Proportion of variance explained
Long term bonds have higher weights in the level factor in high VIX/recession.

**Figure**: Factor Loading to the Level Factor (1st Factor)
Empirical Applications

- In high vix/recession: short term bonds more negative and long term bonds less positive

**Figure:** Factor Loading to the Slope Factor (2nd Factor)

(a) Log-normalized VIX

(b) Recession Indicator
Minimum portfolio weight in the curvature factor shifts to shorter term bond in high vix/recession

Figure: Factor Loading to the Curvature Factor (3rd Factor)
Empirical Applications: Test Constancy of Loadings

- Loadings in low vix are different from loadings in high vix (red region)

**Figure:** Generalized Correlation Test of Estimated Loadings in Two States under Null Hypothesis ($H_0$: Loadings in Two States are Constant)

(a) t-value

(b) p-value
Methodology

- Estimators for latent factors, loadings and common components where loadings are state-dependent
- We combine large dimensional factor modeling with nonparametric estimation
- Asymptotic properties of the estimators
- Constancy test for estimated state-varying factor loadings

Empirical Results

- We discover the movements of factor loadings by state values in the US Treasury Securities and Equity Markets
- Promising empirical results in other data sets
We generate data from a one-factor model

\[ X_{it} = \Lambda_i(S_t)F_t + e_{it} \]

- **Factor:** \( F_t \sim N(0, 1) \)
- **State:** Ornstein–Uhlenbeck (OU) process (mean-reverting)
  \[ S_t = \theta(\mu - S_t)dt + \sigma dW_t, \text{ where } \theta = 1, \mu = 0.2, \text{ and } \sigma = 1 \]
  - stochastic volatility in financial data
- **Loading:** \( \Lambda_i(S_t) = \Lambda_{0i} + \frac{1}{2} S_t \Lambda_{1i} + \frac{1}{4} S_t^2 \Lambda_{2i} + \frac{1}{8} S_t^3 \Lambda_{3i}, \text{ where } \Lambda_{0i}, \Lambda_{1i}, \Lambda_{2i}, \Lambda_{3i} \sim N(0, 1) \)
- **Idiosyncratic errors:** IID/Heteroskedasticity/Cross sectional dependence
Simulation of CLT for Estimated Factors

\[ \sqrt{N} (\hat{\Gamma}_t^s)^{-1/2} (\hat{Q}_s)^{-1} \hat{\Gamma}_s (K_s^{-1/2} (S_t) \hat{F}_t^s - (H_s)' F_t) \xrightarrow{d} N(0, I_r) \]

**Figure**: Comparison between simulated normalized factor distribution and standard normal distribution
Simulation of CLT for Estimated Loadings

\[ \sqrt{T} h(\hat{\Phi}_i^s)^{-1/2}(\hat{Q}^s)'(\hat{\Lambda}_i(s) - (H^s)^{-1}\Lambda_i(s)) \overset{d}{\to} N(0, I_r) \]

**Figure:** Comparison between simulated normalized loading distribution and standard normal distribution
Simulations

Simulation of CLT for Common Component

\[
\left( \frac{1}{N} \hat{V}_{it,s} + \frac{1}{T_h} \hat{W}_{it,s} \right)^{-1/2} \left( \hat{C}_{it,s} - C_{it,s} \right) \xrightarrow{d} N(0, I_r)
\]

**Figure:** Comparison between simulated normalized common component distribution and standard normal distribution
Simulation of CLT for Estimated Generalized Correlation

- Loading: constant with the state $\Lambda_i(S_t) = \Lambda_0i$
- $\sqrt{NTh}(\hat{\rho} - r - \hat{\xi}^T\hat{b})/(\hat{\Omega})^{1/2} \overset{d}{\rightarrow} N(0, 1)$

Figure: Comparison between simulated normalized estimated generalized correlation distribution and standard normal distribution
Recover Functional Form of Loadings vs. State

\[ \Lambda_i(S_t) = \Lambda_{0i} + \frac{1}{2} S_t \Lambda_{1i} + \frac{1}{4} S_t^2 \Lambda_{2i} + \frac{1}{8} S_t^3 \Lambda_{3i} \]

Figure: Loading as a function of the State