# Factor-augmented Regression for High Dimensional Time Series

Dehao Dai (UC San Diego)

CMStatistics, 2023

# High-dimensional Linear Regression

Linear regression model:

$$y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + e_t \text{ or } \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

# High-dimensional Linear Regression

Linear regression model:

$$y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + e_t \text{ or } \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

Challenges:

- Dependence
- Correlation
- Non-Gaussianity
- High-dimensionarity

# High-dimensional Linear Regression

Linear regression model:

$$y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + e_t \text{ or } \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

Correlation:

- PCA: Bai (2003), Fan et al.(2013)
- Sparse regression model
  LASSO: Tibshirani (1996), Huang et al. (2008)
  SCAD: Fan and Li, 2001, Xie and Huang (2009)
- Dantzig Selector: Candes and Tao (2007), Bickel et al. (2009)
- Latent factor model: Bai (2003), Bai and Ng (2003), Fan et al. (2011)

# High-dimensional Linear Regression

Linear regression model:

$$y_t = \boldsymbol{x}_t{}^\top \boldsymbol{\beta} + e_t \text{ or } \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

Our new model:

$$y_t = \boldsymbol{f}_t^\top \boldsymbol{\gamma}^* + \boldsymbol{u}_t^\top \boldsymbol{\beta}^* + e_t$$

$$\boldsymbol{x}_t = \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{u}_t$$

# Factor Model

Model: $x_{it} = \boldsymbol{b}_i^\top \boldsymbol{f}_t + u_{it}$ or $\boldsymbol{X} = \boldsymbol{F}\boldsymbol{B}^\top + \boldsymbol{U}$ and its regular estimation

$$(\hat{\boldsymbol{F}}, \hat{\boldsymbol{B}}) = \arg\min \sum_{i=1}^{p} \sum_{t=1}^{T} (x_{it} - \boldsymbol{b}_i^\top \boldsymbol{f}_t)^2 \text{ subject to } T^{-1}\boldsymbol{F}^\top \boldsymbol{F} = \boldsymbol{I}_K, \boldsymbol{B}^\top \boldsymbol{B} \text{ is diagonal}$$

- Factor: $\hat{\boldsymbol{F}} = \sqrt{T}\hat{\Lambda}$, where $\hat{\Lambda}$ is the matrix of the $K$ eigenvectors corresponding to the first $K$ largest eigenvalues of the $T \times T$ matrix $\boldsymbol{X}\boldsymbol{X}^\top$
- Factor Loading: $\hat{\boldsymbol{B}} = \boldsymbol{X}^\top \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}^\top \hat{\boldsymbol{F}})^{-1} = T^{-1}\boldsymbol{X}^\top \hat{\boldsymbol{F}}$
- Idiosyncratic Error: $\hat{\boldsymbol{U}} = \boldsymbol{X} - \hat{\boldsymbol{F}}\hat{\boldsymbol{B}}^\top = (\boldsymbol{I}_K - T^{-1}\hat{\boldsymbol{F}}\hat{\boldsymbol{F}}^\top)\boldsymbol{X}$

# Factor Model

Model: $x_{it} = \boldsymbol{b}_i^\top \boldsymbol{f}_t + u_{it}$ or $\boldsymbol{X} = \boldsymbol{F}\boldsymbol{B}^\top + \boldsymbol{U}$ and its regular estimation

$$(\hat{\boldsymbol{F}}, \hat{\boldsymbol{B}}) = \arg\min \sum_{i=1}^{p} \sum_{t=1}^{T} (x_{it} - \boldsymbol{b}_i^\top \boldsymbol{f}_t)^2 \text{ subject to } T^{-1}\boldsymbol{F}^\top \boldsymbol{F} = \boldsymbol{I}_K, \boldsymbol{B}^\top \boldsymbol{B} \text{ is diagonal}$$

- Factor: $\hat{\boldsymbol{F}} = \sqrt{T}\hat{\boldsymbol{\Lambda}}$, where $\hat{\boldsymbol{\Lambda}}$ is the matrix of the $K$ eigenvectors corresponding to the first $K$ largest eigenvalues of the $T \times T$ matrix $\boldsymbol{X}\boldsymbol{X}^\top$
- Factor Loading: $\hat{\boldsymbol{B}} = \boldsymbol{X}^\top \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}^\top \hat{\boldsymbol{F}})^{-1} = T^{-1}\boldsymbol{X}^\top \hat{\boldsymbol{F}}$
- Idiosyncratic Error: $\hat{\boldsymbol{U}} = \boldsymbol{X} - \hat{\boldsymbol{F}}\hat{\boldsymbol{B}}^\top = (\boldsymbol{I}_K - T^{-1}\hat{\boldsymbol{F}}\hat{\boldsymbol{F}}^\top)\boldsymbol{X}$

# A Framework of High-dimensional Temporal Data

Weak stationary and causal processes of the nonlinear form

$$\boxed{X_t = (X_{1t}, \ldots, X_{pt})^\top = G(\mathcal{F}_t)}$$

- Input: $\mathcal{F}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$, where $\varepsilon_t$, $t \in \mathbb{Z}$, are i.i.d innovations.
- Function: $G(\cdot) = (g_1(\cdot), \ldots, g_p(\cdot))^\top$.

# Functional Dependence Measures

- Functional dependence measure

$q \geq 2,\ t \geq 0,\ 1 \leq j \leq p$

$$X_t = G(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0, \varepsilon_{-1}, \ldots)$$
$$\Downarrow$$
$$X_t^* = G(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0^*, \varepsilon_{-1}, \ldots)$$

$$\begin{aligned}\delta_{t,q,j} &= \|X_{jt} - X_{jt}^*\|_q \\ &= \|g_j(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0, \varepsilon_{-1}, \ldots) \\ &\quad - g_j(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0^*, \varepsilon_{-1}, \ldots)\|_q\end{aligned}$$

$$\|X_{\cdot j}\|_q = \sup_{m\geq 0} \rho^{-m} \sum_{t=m}^{\infty} \delta_{t,q,j}, \rho \in (0,1)$$

- Dependence-adjusted Orlicz norm

One dimension

$$\|X_\cdot\|_{\psi_\nu} = \sup_{q\geq 2} q^{-\nu} \sum_{t=0}^{\infty} \|X_t - X_t^*\|_q$$

$p$ dimension

$$\|X_\cdot\|_{\psi_\nu} = \sup_{|v|_2\neq 0} \|v^\top X_\cdot\|_{\psi_\nu}/|v|_2$$

# Functional Dependence Measures

- Functional dependence measure

$q \geq 2,\ t \geq 0,\ 1 \leq j \leq p$

$$X_t = G(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0, \varepsilon_{-1}, \ldots)$$
$$\Downarrow$$
$$X_t^* = G(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0^*, \varepsilon_{-1}, \ldots)$$

$$\begin{aligned} \delta_{t,q,j} &= \|X_{jt} - X_{jt}^*\|_q \\ &= \|g_j(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0, \varepsilon_{-1}, \ldots) \\ &\quad - g_j(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0^*, \varepsilon_{-1}, \ldots)\|_q \end{aligned}$$

$$\|X_{\cdot j}\|_q = \sup_{m \geq 0} \rho^{-m} \sum_{t=m}^{\infty} \delta_{t,q,j},\ \rho \in (0,1)$$

- Dependence-adjusted Orlicz norm

One dimension

$$\|X_\cdot\|_{\psi_\nu} = \sup_{q \geq 2} q^{-\nu} \sum_{t=0}^{\infty} \|X_t - X_t^*\|_q$$

$p$ dimension

$$\|X_\cdot\|_{\psi_\nu} = \sup_{|v|_2 \neq 0} \|v^\top X_\cdot\|_{\psi_\nu} / |v|_2$$

# Convergence Rate: Regular Conditions

- Factors: $(\boldsymbol{f}_t)_{t\geq 1}$ is weakly stationary with mean zero
- Factor Loadings: $\boldsymbol{B}$ satisfies
  - $p/\tau \leq \lambda_{\min}\left(\boldsymbol{B}^\top \boldsymbol{B}\right) \leq \lambda_{\max}\left(\boldsymbol{B}^\top \boldsymbol{B}\right) \leq p\tau$ with $\tau > 1$
  - $|\boldsymbol{B}|_{\max} \leq c$ with $c > 0$
- Time and Cross-Section Dependence and Heteroskedasticity:
  - $(\boldsymbol{u}_t)_{t\geq 1}$ is weakly stationary with mean zero
  - $c_1 < \lambda_{\min}(\boldsymbol{\Sigma}_u)$, $|\boldsymbol{\Sigma}_u|_1 < c_2$ and $\min_{i,j\leq p} \mathrm{Var}\left(u_{it}u_{jt}\right) > c_1$ with $c_1 < c_2$
  - $\mathbb{E}|\boldsymbol{u}_s^\top \boldsymbol{u}_t - E\boldsymbol{u}_s^\top \boldsymbol{u}_t|^4 < cp^2$ with $c > 0$
- Uncorrelation between factors and idiosyncratic errors: $\mathbb{E}u_{it}f_{jt} = 0$
- $\mathbb{E}|\sum_{i=1}^p \boldsymbol{b}_i u_{it}|_2^4 < cp^2$ with $c > 0$

# Convergence Rate: Temporal Dependence

$(\boldsymbol{u}_t, \boldsymbol{f}_t)_{t\geq 1}$ satisfies

- $(\boldsymbol{u}_t, \boldsymbol{f}_t)_{t\geq 1} \in \mathcal{L}^q$ for all $q > 2$
- Let

$$\|\boldsymbol{u}_{\cdot j}\|_{\psi_\mu} := \sup_{q\geq 2} q^{-\mu}\|\boldsymbol{u}_{\cdot j}\|_q, \text{ and } \|\boldsymbol{f}_{\cdot j}\|_{\psi_\nu} := \sup_{q\geq 2} q^{-\nu}\|\boldsymbol{f}_{\cdot j}\|_q$$

and

$$\|\boldsymbol{u}_{\cdot}\|_{\psi_\mu} = \max_{1\leq j\leq p} \|\boldsymbol{u}_{\cdot j}\|_{\psi_\mu} \text{ and } \|\boldsymbol{f}_{\cdot}\|_{\psi_\nu} = \max_{1\leq j\leq p} \|\boldsymbol{f}_{\cdot j}\|_{\psi_\nu}$$

$\|\boldsymbol{u}_{\cdot}\|_{\psi_\mu}, \|\boldsymbol{f}_{\cdot}\|_{\psi_\nu} < \infty$ for $0 < \rho < 1$ and some $\mu, \nu > 0$

# Convergence Rate

Let

$$\boldsymbol{H} = T^{-1}\boldsymbol{V}^{-1}\hat{\boldsymbol{F}}^\top \boldsymbol{F}\boldsymbol{B}^\top \boldsymbol{B},$$

where $\boldsymbol{V} \in \mathbb{R}^{K\times K}$ is a diagonal matrix consisting of the first $K$ largest eigenvalues of the matrix $T^{-1}\boldsymbol{X}\boldsymbol{X}^\top$. Then, we have

- $|\hat{\boldsymbol{F}} - \boldsymbol{F}\boldsymbol{H}^\top|_{\mathbb{F}} = O_{\mathbb{P}}\left(\sqrt{\frac{T}{p}} + \frac{1}{\sqrt{T}}\right)$.
- $|\boldsymbol{H}^\top\boldsymbol{H} - \boldsymbol{I}_K|_{\mathbb{F}} = O_{\mathbb{P}}\left(\sqrt{\frac{1}{T}} + \sqrt{\frac{1}{p}}\right)$.
- $\max_{j\le p} |\hat{\boldsymbol{b}}_j - \boldsymbol{H}\boldsymbol{b}_j|_2 = O_{\mathbb{P}}\left(\sqrt{1/p} + (\log p)^{1/2+\mu+\nu}/\sqrt{T}\right)$.
- For any $\mathcal{I} \subset \{1, 2, \ldots, p\}$, we have

$$\max_{j\in\mathcal{I}} \sum_{t=1}^{T} \left|\hat{u}_{tj} - u_{tj}\right|^2 = O_{\mathbb{P}}\left((\log|\mathcal{I}|)^{1+2\mu+2\nu} + T/p\right).$$

# Lasso Estimation

- Motivation:
  - High dimension setting: sample size $T <$ dimension $p$
  - Assume the true parameter $\boldsymbol{\beta}$ is sparse
- Regularization:

$$(\hat{\boldsymbol{\beta}}_\lambda, \hat{\boldsymbol{\gamma}}) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{\gamma}\in\mathbb{R}^K}{\arg\min} \left\{ \frac{1}{2T} |\boldsymbol{Y} - \hat{\boldsymbol{U}}\boldsymbol{\beta} - \hat{\boldsymbol{F}}\boldsymbol{\gamma}|_2^2 + \lambda |\boldsymbol{\beta}|_1 \right\},$$

  where $\lambda > 0$ controls the strength of regularization.
- Equivalence:

$$\begin{aligned} \tilde{\boldsymbol{Y}} &= (\boldsymbol{I} - \hat{\boldsymbol{P}})\boldsymbol{Y} \\ &= (\boldsymbol{I} - \tfrac{1}{T}\hat{\boldsymbol{F}}\hat{\boldsymbol{F}}^\top)\boldsymbol{Y} \end{aligned} \quad \Rightarrow \quad \begin{aligned} \hat{\boldsymbol{\beta}}_\lambda &= \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \left\{ \frac{1}{2T} |\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{U}}\boldsymbol{\beta}|_2^2 + \lambda |\boldsymbol{\beta}|_1 \right\} \\ \hat{\boldsymbol{\gamma}} &= (\hat{\boldsymbol{F}}^\top \hat{\boldsymbol{F}})^{-1} \hat{\boldsymbol{F}}^\top \boldsymbol{Y} = \frac{1}{T} \hat{\boldsymbol{F}}^\top \boldsymbol{Y} \end{aligned}$$

# Convergence Rate

$(e_t)_{t\geq 1}$ satisfies

- weakly stationary with mean 0 and uncorrelated with $\boldsymbol{u}_t$ and $\boldsymbol{f}_t$,
- $\|e_.\|_q := \sup_{m\geq 0} \rho^{-m}\Psi_{m,q} < \infty$ with $\Psi_{m,q} = \sum_{i=m}^{\infty} \psi_{i,q}$ for some $0 < \rho < 1$,
- $\|e_.\|_{\psi_\gamma} := \sup_{q\geq 2} q^{-\gamma}\|e_.\|_q < \infty$. for some $\gamma \geq 0$.

# Convergence Rate

$(e_t)_{t\geq 1}$ satisfies

- weakly stationary with mean 0 and uncorrelated with $\boldsymbol{u}_t$ and $\boldsymbol{f}_t$,
- $\|e_\cdot\|_q := \sup_{m\geq 0} \rho^{-m} \Psi_{m,q} < \infty$ with $\Psi_{m,q} = \sum_{i=m}^{\infty} \psi_{i,q}$ for some $0 < \rho < 1$,
- $\|e_\cdot\|_{\psi_\gamma} := \sup_{q\geq 2} q^{-\gamma} \|e_\cdot\|_q < \infty$. for some $\gamma \geq 0$.

If $\boldsymbol{\varphi}^* = \boldsymbol{\gamma}^* - \boldsymbol{B}^\top \boldsymbol{\beta}^*$, then we have

$$|\hat{\boldsymbol{\gamma}} - \boldsymbol{H}\boldsymbol{\gamma}^*|_2 = O_{\mathbb{P}} \left\{ \frac{1}{\sqrt{T}} + \left( \frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p}} \right) |\boldsymbol{\varphi}^*|_2 + \left( \frac{(\log |\mathcal{S}_\star|)^{\frac{1}{2}+\nu+\mu}}{\sqrt{T}} + \frac{1}{\sqrt{p}} \right) |\boldsymbol{\beta}^*|_1 \right\}$$

where $\mathcal{S}_* = \{j : \beta_j^* \neq 0, 1 \leq j \leq p\}$ and $|\mathcal{S}_*|$ is cardinality of set $\mathcal{S}_*$.

# Convergence Rate

$(e_t)_{t\geq 1}$ satisfies

- weakly stationary with mean 0 and uncorrelated with $\boldsymbol{u}_t$ and $\boldsymbol{f}_t$,
- $\|e_\cdot\|_q := \sup_{m\geq 0} \rho^{-m}\Psi_{m,q} < \infty$ with $\Psi_{m,q} = \sum_{i=m}^{\infty} \psi_{i,q}$ for some $0 < \rho < 1$,
- $\|e_\cdot\|_{\psi_\gamma} := \sup_{q\geq 2} q^{-\gamma}\|e_\cdot\|_q < \infty$. for some $\gamma \geq 0$.

If

$$|\mathcal{S}|\left(\frac{1}{p} + \frac{(\log p)^{1+2\mu+2\nu}}{T}\right) \to 0,$$

then choosing appropriate $\lambda = \frac{c}{T}|\hat{\boldsymbol{U}}^\top(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{U}}\boldsymbol{\beta}^*)|_\infty$ with some constant $c \geq 2$, we have $\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^* \in \mathcal{C}(\mathcal{S}_*, 3)$ and

$$|\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*|_2 = O_{\mathbb{P}}\left(\sqrt{\frac{|\mathcal{S}_*|(\log p)^{1+2\gamma+2\mu}}{T}} + \frac{\mathcal{V}_{T,p}|\boldsymbol{\varphi}^*|_2\sqrt{|\mathcal{S}_*|}}{T}\right),$$

$$|\hat{\boldsymbol{U}}(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*)|_2^2 = O_{\mathbb{P}}\left(|\mathcal{S}_*|(\log p)^{1+2\gamma+2\mu} + \frac{\mathcal{V}_{T,p}^2|\boldsymbol{\varphi}^*|_2^2|\mathcal{S}_*|}{T}\right)$$

with

$$\mathcal{V}_{T,p} = \frac{T}{p} + \sqrt{\frac{(\log p)^{1+2\mu+2\nu}}{T}} + \sqrt{\frac{T(\log p)^{1+2\mu+2\max\{\nu,\mu\}}}{p}}.$$

# Debiased Lasso Estimator

Let $\hat{\mathbf{\Theta}}$ be an approximation for the inverse of $\tilde{\mathbf{\Sigma}}_{\boldsymbol{u}} = \frac{1}{T}\hat{\boldsymbol{U}}^\top \hat{\boldsymbol{U}}$, we get the error between de-biased estimator $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ by

$$
\begin{aligned}
\tilde{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^* =& \hat{\boldsymbol{\beta}}_\lambda + \frac{1}{T}\hat{\mathbf{\Theta}}\hat{\boldsymbol{U}}^\top \left(\boldsymbol{Y} - \hat{\boldsymbol{U}}\hat{\boldsymbol{\beta}}_\lambda\right) - \boldsymbol{\beta}^* \\
=& \frac{1}{T}\hat{\mathbf{\Theta}}\hat{\boldsymbol{U}}^\top \boldsymbol{e} + \frac{1}{T}\hat{\mathbf{\Theta}}\hat{\boldsymbol{U}}^\top \boldsymbol{F}\boldsymbol{\varphi}^* + (\boldsymbol{I}_p - \hat{\mathbf{\Theta}}\tilde{\mathbf{\Sigma}}_{\boldsymbol{u}})(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*).
\end{aligned}
$$

# Debiased Lasso Estimator

Let $\hat{\mathbf{\Theta}}$ be an approximation for the inverse of $\tilde{\mathbf{\Sigma}}_{\boldsymbol{u}} = \frac{1}{T}\hat{\boldsymbol{U}}^\top \hat{\boldsymbol{U}}$, we get the error between de-biased estimator $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ by

$$
\begin{aligned}
\tilde{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^* &= \hat{\boldsymbol{\beta}}_\lambda + \frac{1}{T}\hat{\mathbf{\Theta}}\hat{\boldsymbol{U}}^\top \left(\boldsymbol{Y} - \hat{\boldsymbol{U}}\hat{\boldsymbol{\beta}}_\lambda\right) - \boldsymbol{\beta}^* \\
&= \frac{1}{T}\hat{\mathbf{\Theta}}\hat{\boldsymbol{U}}^\top \boldsymbol{e} + \frac{1}{T}\hat{\mathbf{\Theta}}\hat{\boldsymbol{U}}^\top \boldsymbol{F}\boldsymbol{\varphi}^* + (\boldsymbol{I}_p - \hat{\mathbf{\Theta}}\tilde{\mathbf{\Sigma}}_{\boldsymbol{u}})(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*).
\end{aligned}
$$

## Gaussian Approximation

Assume that $e_t$ satisfies the regular conditions and temporal dependence, and $\min_{1\le j\le p} \sigma_{jj} \ge b_0$ for some universal constant $b_0 > 0$. Then, we have

$$
d_n := \sup_{x\ge 0}\left|\mathbb{P}\left(\frac{1}{\sqrt{T}}|\mathbf{\Theta}\boldsymbol{U}^\top \boldsymbol{e}|_\infty \le x\right) - \mathbb{P}(|\boldsymbol{Z}|_\infty \le x)\right| \lesssim |\mathbf{\Theta}|_\infty \|\boldsymbol{u}_\cdot\|_{\psi_\mu} \|e_\cdot\|_{\psi_\gamma}\left(\frac{\log^c(p\vee T)}{T}\right)^{\frac{1}{9}},
$$

with $c = \max\{4\mu + 4\gamma + 7, 2\mu + 2\gamma + 10\}$, and the constant in $\lesssim$ depends on $\mu$, $\gamma$, $\rho$ and $b_0$.

# Debiased Lasso Estimator

Let $\hat{\mathbf{\Theta}}$ be an approximation for the inverse of $\tilde{\mathbf{\Sigma}}_{\boldsymbol{u}} = \frac{1}{T}\hat{\boldsymbol{U}}^{\top}\hat{\boldsymbol{U}}$, we get the error between de-biased estimator $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ by

$$
\begin{aligned}
\tilde{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^* &= \hat{\boldsymbol{\beta}}_\lambda + \frac{1}{T}\hat{\mathbf{\Theta}}\hat{\boldsymbol{U}}^{\top}\left(\boldsymbol{Y} - \hat{\boldsymbol{U}}\hat{\boldsymbol{\beta}}_\lambda\right) - \boldsymbol{\beta}^* \\
&= \frac{1}{T}\hat{\mathbf{\Theta}}\hat{\boldsymbol{U}}^{\top}\boldsymbol{e} + \frac{1}{T}\hat{\mathbf{\Theta}}\hat{\boldsymbol{U}}^{\top}\boldsymbol{F}\boldsymbol{\varphi}^* + (\boldsymbol{I}_p - \hat{\mathbf{\Theta}}\tilde{\mathbf{\Sigma}}_{\boldsymbol{u}})(\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*).
\end{aligned}
$$

## Gaussian Multiplier

Assume that

$$
(\lambda_{\max}|\mathcal{S}_*| + \Delta_\infty)(\log p)^{1+\mu+\gamma} \to 0
$$

$$
\left(\mathcal{V}_{T,p}|\boldsymbol{\varphi}^*|_2 + \sqrt{\frac{T}{p} + (\log p)^{1+\mu+\gamma}}\right)|\mathbf{\Theta}|_\infty \sqrt{\frac{(\log p)^{1+2\mu+2\gamma}}{T}} \to 0
$$

and $\min_{1\le j\le p}\sigma_{jj} \ge b_0$ for some universal constant $b_0 > 0$. Then, we have

$$
\sup_{x\ge 0}\left|\mathbb{P}\left(\sqrt{T}|\tilde{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*|_\infty \le x\right) - \mathbb{P}(|\boldsymbol{Z}|_\infty \le x)\right| \to 0
$$

# Estimation of Unknown Long-run Covariance Matrices

Covariance matrix of $\boldsymbol{Z}$: $\boldsymbol{\Sigma}_{\boldsymbol{Z}} = \mathrm{Cov}(\boldsymbol{Z}) = \boldsymbol{\Theta}\boldsymbol{\Sigma}_{\boldsymbol{u}e}\boldsymbol{\Theta}^\top$

- Estimate $\boldsymbol{\Theta}$ by node-wise regression in Van de Geer et al. (2014)

$$\hat{\boldsymbol{w}}_j = \underset{\boldsymbol{w}\in\mathbb{R}^{p-1}}{\arg\min}\left\{\frac{1}{2T}\sum_{t=1}^{T}|\hat{u}_{tj} - \boldsymbol{w}^\top\hat{\boldsymbol{u}}_{t,-j}|^2 + \lambda_j|\boldsymbol{w}|_1\right\}$$

  and $\hat{\boldsymbol{\Theta}}$ with $\hat{\Theta}_{jl} = -\hat{w}_{jl}/\hat{v}_j^2$ for all $j \neq l$ and $\hat{\Theta}_{jj} = -1/\hat{v}_j^2$ with

$$\hat{v}_j^2 = \frac{1}{T}\sum_{t=1}^{T}|\hat{u}_{tj} - \hat{\boldsymbol{w}}_j^\top\hat{\boldsymbol{u}}_{t,-j}|^2 + \lambda_j|\hat{\boldsymbol{w}}_j|_1.$$

- Estimate $\boldsymbol{\Sigma}_{\boldsymbol{u}e} := (\sigma_{jk})_{j,k=1}^{p} := \sum_{\ell=-\infty}^{\infty}\Gamma_{\boldsymbol{u}}(\ell)\gamma_e(\ell)$ by

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}e} = \frac{1}{Mw}\sum_{b=1}^{w}\left(\sum_{t\in L_b}\hat{\boldsymbol{u}}_t\hat{e}_t - M\bar{\boldsymbol{Q}}\right)\left(\sum_{t\in L_b}\hat{\boldsymbol{u}}_t\hat{e}_t - M\bar{\boldsymbol{Q}}\right)^\top$$

  where $\bar{\boldsymbol{Q}} = (Mw)^{-1}\sum_{t=1}^{Mw}\hat{\boldsymbol{u}}_t\hat{e}_t$, the $b$-th window $L_b = \{1+(b-1)M,\ldots,bM\}$, $b = 1,2,\ldots,w$, the window size $|L_b| = M \to \infty$ and the number of blocks $w = \lfloor T/M \rfloor$.

# Convergence of Covariance Estimate

Assume that $e_t$ satisfies the regular conditions and temporal dependence, and $\min_{1\le j\le p}\sigma_{jj} \ge b_0$ for some universal constant $b_0 > 0$. Then, we have

$$\mathbb{P}(T|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}e} - \boldsymbol{\Sigma}_{\boldsymbol{u}e}|_\infty \ge x) \lesssim p^2 \exp\left(-\frac{x^c}{4ec(\sqrt{w}\|\boldsymbol{u}_\cdot\|_{\psi_\mu}^2\|e_\cdot\|_{\psi_\gamma}^2)^c}\right),$$

with $c = 1/(1 + 2\mu + 2\gamma)$ and the constants in $\lesssim$ only depend on $\mu$ and $\gamma$.

# Hypothesis test

$$H_0 : \boldsymbol{\beta}^* = 0 \text{ versus } H_1 : \boldsymbol{\beta}^* \neq 0.$$

- Given the level $\alpha \in (0, 1)$, we can reject $H_0$ if $\sqrt{T}|\tilde{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}_0|_\infty \geq \hat{\chi}_{1-\alpha}$
- $(1-\alpha)$th confidence intervals for $\boldsymbol{\beta}^*$ can be constructed as $\tilde{\boldsymbol{\beta}}_\lambda \pm \hat{\chi}_{1-\alpha}/\sqrt{T}$.

## Theorem

Assume $|\boldsymbol{I}_p - \hat{\boldsymbol{\Theta}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{u}}|_{\max} = O_{\mathbb{P}}(\Lambda_{\max}), |\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}|_\infty = O_{\mathbb{P}}(\Delta_\infty)$, and $|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}|_{\max} = O_{\mathbb{P}}(\Delta_{\max})$. Further assume that

$$\Delta_{\max}^2 \frac{(\log p)^{1+2\mu+2\gamma}}{T} + \Delta_\infty^2 + \Delta_\infty|\boldsymbol{\Theta}|_\infty = o\left(\frac{1}{\log p}\right)$$

Then

$$\sup_{x \geq 0} \left|\mathbb{P}(\sqrt{T}|\tilde{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*|_\infty \leq x) - \mathbb{P}^*(|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{Z}}^{1/2}\boldsymbol{\xi}|_\infty \leq x)\right| \xrightarrow{\mathbb{P}} 0$$

# Simulation Study: Setting A

- $\boldsymbol{\gamma}^* = (0.5, 0.5)$, $K = 2$
- $\boldsymbol{\beta}^* = (0.5, 0.5, 0.5, 0, \ldots, 0)$, dimension $p = 100$ and sparsity $s = 3$
- $\boldsymbol{f}_t \sim \mathcal{N}(0, \boldsymbol{I}_K)$
- $\boldsymbol{u}_t \sim \mathcal{N}(0, \boldsymbol{I}_p)$
- $\boldsymbol{B}$ ~Unif$(-1, 1)$
- $e_t = 0.1e_{t-1} + \eta_t$ with innovation $\eta_t$ i.i.d from
  - Gaussian distribution $\mathcal{N}(0, 0.5^2)$, $\gamma = 1/2$
  - double exponential distribution (Laplace$(0, \frac{1}{2\sqrt{2}})$), $\gamma = 1$
- Sample size $T$
  - sub-Gaussian tail: $s\sqrt{(\log p)^{1+2\gamma}/T} \in [0.30, 0.60]$
  - sub-Exponential tail: $s\sqrt{(\log p)^{1+2\gamma}/T} \in [0.50, 0.80]$

# Simulation Study: Result

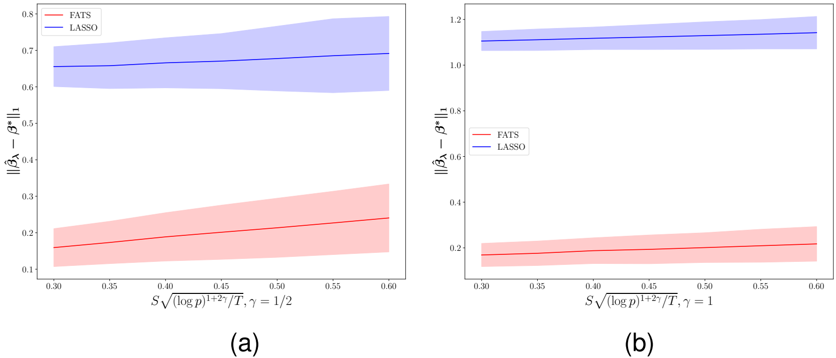



Figure: Accuracy for with $|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*|_1$ based on 500 replications with different innovations in noise $\boldsymbol{e}$

# Simulation Study: Setting B

- $\boldsymbol{\gamma}^* = (0.5, 0.5)$, $K = 2$
- $\boldsymbol{\beta}^* = (0.5, 0.5, 0.5, 0, \ldots, 0)$, dimension $p = 100$ and sparsity $s = 3$
- $t$-th row of $\boldsymbol{F}$ and $\boldsymbol{U}$ satisfy
  - $\boldsymbol{f}_t = \boldsymbol{\Phi} \boldsymbol{f}_{t-1} + \boldsymbol{\epsilon}_t$ with $\boldsymbol{\Phi}_{i,j} = 0.1^{|i-j|+1}$, $1 \le i, j \le K$ and $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{I}_K)$; $\boldsymbol{u}_t \sim \mathcal{N}(0, \boldsymbol{I}_p)$
  - $\boldsymbol{u}_t = \boldsymbol{\Phi} \boldsymbol{u}_{t-1} + \boldsymbol{\varepsilon}_t$ with $\boldsymbol{\Phi}_{i,j} = 0.1^{|i-j|+1}$, $1 \le i, j \le p$ and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \boldsymbol{I}_p)$; $\boldsymbol{f}_t \sim \mathcal{N}(0, \boldsymbol{I}_K)$
- $\boldsymbol{e} \sim \mathcal{N}(0, 0.5^2)$
- $\boldsymbol{B}$ ~Unif$(-1, 1)$
- Sample size $T$:
  - $\boldsymbol{F}$ with VAR(1): $s\sqrt{\log p/T} \in [0.30, 0.60]$
  - $\boldsymbol{U}$ with VAR(1): $s\sqrt{(\log p)^{1+2\mu}/T} \in [0.30, 0.60]$

# Simulation Study: Result

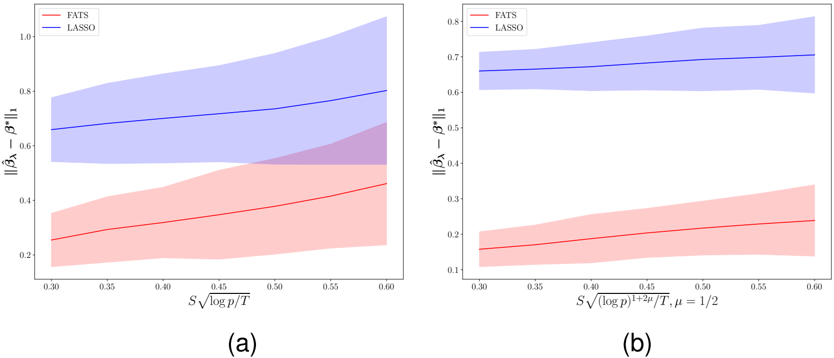



Figure: Accuracy for with $|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*|_1$ based on 500 replications with VAR(1) $\boldsymbol{F}$ and $\boldsymbol{U}$.

# Simulation Study: Gaussian Approximation

Settings:

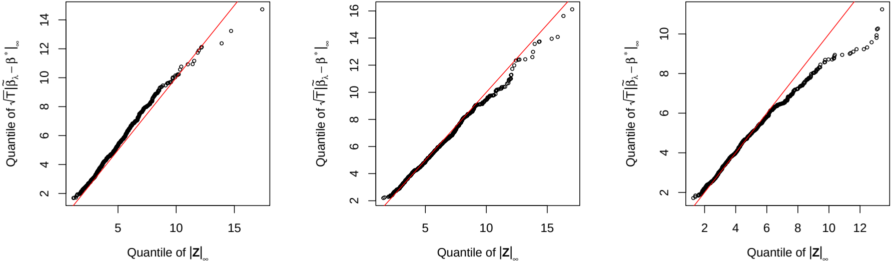
- $\boldsymbol{\gamma}^* = (0.5, 0.5)^\top$, $K = 2$
- $\boldsymbol{\beta}^* = (0.5, 0.5, 0.5, 0, \ldots, 0)^\top$ with dimension $p = 100$ and sparsity $s = 3$
- $\boldsymbol{f}_t \sim \mathcal{N}(0, \boldsymbol{I}_K)$
- $\boldsymbol{u}_t \sim \mathcal{N}(0, \boldsymbol{I}_p)$
- $\boldsymbol{B} \sim$ Unif $(-1, 1)$
- $e_t = \phi e_{t-1} + \eta_t = \sum_{i=0}^{\infty} \phi^i \eta_{t-i}$ with (1) $\phi = 0.1$; (2) $\phi = 0.5$; (3) $\phi = 0.9$
- Sample size $T = 300$

# Simulation Study: Gaussian Approximation

Implementation:

- Estimate factor $\hat{\boldsymbol{F}}$, loading matrix $\hat{\boldsymbol{B}}$, and noise $\hat{\boldsymbol{U}}$.
- For the given $\hat{\boldsymbol{U}}$, we estimate $\hat{\boldsymbol{\Theta}}$ by using node-wise regression.
- Use the Lasso method to estimate $\boldsymbol{\beta}_\lambda$ based on data $\hat{\boldsymbol{U}}$ and $\tilde{\boldsymbol{Y}}$ with $\lambda$ chosen by cross-validation.
- Construct a de-biased estimator $\tilde{\boldsymbol{\beta}}_\lambda = \hat{\boldsymbol{\beta}}_\lambda + \frac{1}{T}\hat{\boldsymbol{\Theta}}\hat{\boldsymbol{U}}^\top\left(\boldsymbol{Y} - \hat{\boldsymbol{U}}\hat{\boldsymbol{\beta}}_\lambda\right)$.
- Construct the batched mean estimate $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{Z}}$ from blocks $\hat{\boldsymbol{\Theta}}\hat{\boldsymbol{U}}^\top\hat{e}$ and window size $w = T/M$ with block size $M = 10$.
- Obtain $\boldsymbol{Z} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{Z}}^{1/2}\boldsymbol{\xi}$ with every entry of $\boldsymbol{\xi}$ simulated from $\mathcal{N}(0,1)$ and repeat this computation 1000 times to obtain the $|\boldsymbol{Z}|_\infty$.

# Simulation Study: Gaussian Approximation




(a) $\phi = 0.1$ (b) $\phi = 0.5$ (c) $\phi = 0.9$

Figure: The QQ-plot of $\sqrt{T}|\tilde{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*|_\infty$ versus $|Z|_\infty$ under different dependence $\phi$

# Real Data Study

- Dataset:
  - **IPNCONGD**: industrial production (IP) index with non-durable consumer goods, measuring the real output of all products with few use times located in the United States
  - **HOUSTW**: new privately-owned housing units in the West Census region including California
- Quarter Data from March 1967 to December 2022
- Transformation code: 1 = no transformation, 2 = first difference $\Delta x_t$, 3 = second difference $\Delta^2 x_t$, 4 = $\log(x_t)$, 5 = first difference of logged variables $\Delta \log(x_t)$, 6 = second difference of logged variables $\Delta^2 \log(x_t)$

# Real Data Study

- Evaluation:
  - Given a window size $w$, we use $w$ observation pairs $\{(\boldsymbol{x}_{t-w}, Y_{t-w}), \ldots, (\boldsymbol{x}_{t-1}, Y_{t-1})\}$
  - Output: a prediction $\widehat{Y}_t$ and in-sample mean $\bar{Y}_t = \frac{1}{w}\sum_{i=t-w}^{t-1} Y_i$
  - Out-of-sample $R^2$ is given by

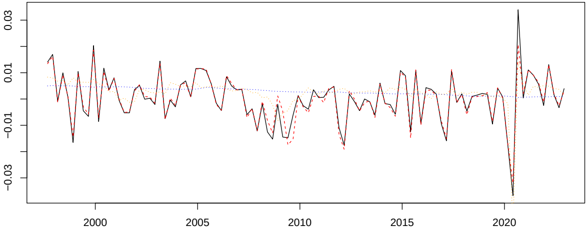
$$R^2 := 1 - \frac{\sum_{t=w+1}^{T}(Y_t - \widehat{Y}_t)^2}{\sum_{t=w+1}^{T}(Y_t - \bar{Y}_t)^2},$$

- Time window $w = 120$ quarterly from March 1967 to December

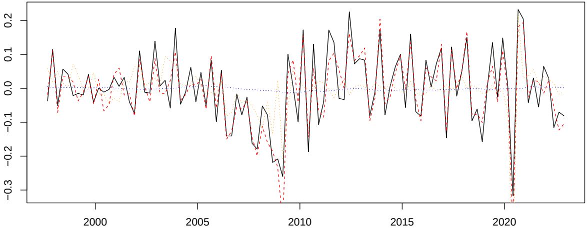
| Data | FATS | PCR |
|---|---|---|
| IPNCONGD | 0.948 | 0.313 |
| HOUSTW | 0.842 | 0.180 |

# Real Data Study: September 1997 to December 2022

- IPNCONG




- HOUSTW



# Reference


J. Bai. (2003). Inferential theory for factor models of large dimensions. Econometrica.

J. Bai and S. Ng. (2002). Determining the number of factors in approximate factor models. Econometrica.

V. Chernozhukov, D. Chetverikov, and K. Kato. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. AOS.

V. Chernozhukov, D. Chetverikov, and K. Kato. (2017). Central limit theorems and bootstrap in high dimensions. AOP.

J. Fan, Y. Liao, and M. Mincheva. (2011). High dimensional covariance matrix estimation in approximate factor models. AOS.

J. Fan, Y. Liao, and M. Mincheva. (2013). Large covariance estimation by thresholding principal orthogonal complements. JRSSB.

W. Liu and W. B. Wu. (2010). Asymptotics of spectral density estimates. Econometric Theory.

M. McCracken and S. Ng. (2020). Fred-qd: A quarterly database for macroeconomic research. Technical report, National Bureau of Economic Research.

S. Van de Geer, P. B uhlmann, Y. Ritov, and R. Dezeure. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. AOS.

D. Zhang and W. B. Wu. (2017). Gaussian approximation for high dimensional time series. AOS.

Thank you for your attention today!