Fixed T Estimation of Linear Panel Data Models with Interactive Fixed Effects

Ayden Higgins

August 22nd 2022

Model

- In a panel with entries indexed $i = 1, \ldots, n$, $t = 1, \ldots, T$, outcomes are generated according to

$$\boldsymbol{y}_t = \alpha \boldsymbol{y}_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\eta}_t,$$

where

$$\boldsymbol{\eta}_t = \Lambda^* \boldsymbol{f}_t^* + \boldsymbol{\varepsilon}_t.$$

Model

- In a panel with entries indexed $i = 1, \ldots, n$, $t = 1, \ldots, T$, outcomes are generated according to

$$\boldsymbol{y}_t = \alpha \boldsymbol{y}_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\eta}_t,$$

where

$$\boldsymbol{\eta}_t = \Lambda^* \boldsymbol{f}_t^* + \boldsymbol{\varepsilon}_t.$$

- Parameter of interest $\boldsymbol{\theta} \coloneqq (\alpha, \boldsymbol{\beta}^\top)^\top$.

Model

- In a panel with entries indexed $i = 1, \ldots, n$, $t = 1, \ldots, T$, outcomes are generated according to

$$\boldsymbol{y}_t = \alpha \boldsymbol{y}_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\eta}_t,$$

where

$$\boldsymbol{\eta}_t = \Lambda^* \boldsymbol{f}_t^* + \boldsymbol{\varepsilon}_t.$$

- Parameter of interest $\boldsymbol{\theta} \coloneqq (\alpha, \boldsymbol{\beta}^\top)^\top$.
- Generalisation of classic models such as individual, time or group effects:

$$\Lambda^* = \begin{pmatrix} \lambda_1 & 1 \\ \vdots & \vdots \\ \lambda_n & 1 \end{pmatrix}, \ \boldsymbol{f}_t^* = \begin{pmatrix} 1 \\ f_t \end{pmatrix}.$$

Model

- In a panel with entries indexed $i = 1, \dots, n$, $t = 1, \dots, T$, outcomes are generated according to

$$\boldsymbol{y}_t = \alpha \boldsymbol{y}_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\eta}_t,$$

where

$$\boldsymbol{\eta}_t = \Lambda^* \boldsymbol{f}_t^* + \boldsymbol{\varepsilon}_t.$$

- Parameter of interest $\boldsymbol{\theta} \coloneqq (\alpha, \boldsymbol{\beta}^\top)^\top$.
- Generalisation of classic models such as individual, time or group effects:

$$\Lambda^* = \begin{pmatrix} \lambda_1 & 1 \\ \vdots & \vdots \\ \lambda_n & 1 \end{pmatrix}, \ \boldsymbol{f}_t^* = \begin{pmatrix} 1 \\ f_t \end{pmatrix}.$$

- $\Lambda^* \boldsymbol{F}^{*\top}$ is low rank.

Problem

- Multiple estimation approaches:

Problem

- Multiple estimation approaches:
  - LS-IFE Estimator: Bai (2009); Moon and Weidner (2015,2017).
    - Easy to implement & General.
    - Consistent as $n, T \to \infty$, inconsistent with fixed T.
    - Biases: (1) Dynamics (2) Cross-sectional dep./hetero. (3) Temporal dep./hetero.

Problem

- Multiple estimation approaches:
  - LS-IFE Estimator: Bai (2009); Moon and Weidner (2015,2017).
    - Easy to implement & General.
    - Consistent as $n, T \to \infty$, inconsistent with fixed T.
    - Biases: (1) Dynamics (2) Cross-sectional dep./hetero. (3) Temporal dep./hetero.
  - CCE Approaches: Pesaran (2006); Bai and Li (2014); Westerlund and Urbain (2015); Westerlund (2020); ...
    - Assumes $\boldsymbol{x}_t = \Gamma \boldsymbol{f}_t + \boldsymbol{e}_t$.
    - Static: consistent with fixed T or $n, T \to \infty$.
    - Dynamic: consistent with $n, T \to \infty$.

Problem

- Multiple estimation approaches:
  - LS-IFE Estimator: Bai (2009); Moon and Weidner (2015,2017).
    - Easy to implement & General.
    - Consistent as $n, T \to \infty$, inconsistent with fixed T.
    - Biases: (1) Dynamics (2) Cross-sectional dep./hetero. (3) Temporal dep./hetero.
  - CCE Approaches: Pesaran (2006); Bai and Li (2014); Westerlund and Urbain (2015); Westerlund (2020); ...
    - Assumes $\boldsymbol{x}_t = \Gamma \boldsymbol{f}_t + \boldsymbol{e}_t$.
    - Static: consistent with fixed T or $n, T \to \infty$.
    - Dynamic: consistent with $n, T \to \infty$.
  - Quasi-difference Approaches: Holtz-Eakin, Newey, Rosen (1988); Ahn et al. (2001); Ahn et al. (2013).
    - Static & Dynamic: consistent with fixed T or $n, T \to \infty$.
    - Difficult to implement.
    - Identification failure.

Problem

- Multiple estimation approaches:
  - LS-IFE Estimator: Bai (2009); Moon and Weidner (2015,2017).
    - Easy to implement & General.
    - Consistent as $n, T \to \infty$, inconsistent with fixed T.
    - Biases: (1) Dynamics (2) Cross-sectional dep./hetero. (3) Temporal dep./hetero.
  - CCE Approaches: Pesaran (2006); Bai and Li (2014); Westerlund and Urbain (2015); Westerlund (2020); ...
    - Assumes $\boldsymbol{x}_t = \Gamma \boldsymbol{f}_t + \boldsymbol{e}_t$.
    - Static: consistent with fixed T or $n, T \to \infty$.
    - Dynamic: consistent with $n, T \to \infty$.
  - Quasi-difference Approaches: Holtz-Eakin, Newey, Rosen (1988); Ahn et al. (2001); Ahn et al. (2013).
    - Static & Dynamic: consistent with fixed T or $n, T \to \infty$.
    - Difficult to implement.
    - Identification failure.
  - Others: Bai (2013); Hsiao et al. (2021); ...

Proposal

- Transformation to reduce the dimension of the model.

Proposal

- Transformation to reduce the dimension of the model.
- Enables the LS-IFE estimator to produce consistent with T fixed:

Proposal

- Transformation to reduce the dimension of the model.
- Enables the LS-IFE estimator to produce consistent with T fixed:
  - Asymptotically unbiased with T fixed even with dynamics, cross-sectional and temporal dependence and/or heteroskedasticity.

Proposal

- Transformation to reduce the dimension of the model.
- Enables the LS-IFE estimator to produce consistent with T fixed:
  - Asymptotically unbiased with T fixed even with dynamics, cross-sectional and temporal dependence and/or heteroskedasticity.
- **Does not remove incidental parameters.**

Outline I: Transformation

- Original model in matrices:

$$\begin{aligned}\boldsymbol{Y}\boldsymbol{S}(\alpha) &= \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k + \boldsymbol{y}_0\boldsymbol{s}^\top(\alpha) + \Lambda^*\boldsymbol{F}^{*\top} + \varepsilon,\\ &= \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k + \Lambda\boldsymbol{F}^\top + \varepsilon.\end{aligned}$$

Outline I: Transformation

- Original model in matrices:

$$\boldsymbol{Y}\boldsymbol{S}(\alpha) = \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k + \boldsymbol{y}_0 \boldsymbol{s}^\top(\alpha) + \Lambda^* \boldsymbol{F}^{*\top} + \varepsilon,$$
$$= \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k + \Lambda \boldsymbol{F}^\top + \varepsilon.$$

- Define the $n \times TK$ matrix $\mathfrak{X} := (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K)$, and $\boldsymbol{Q} := \mathfrak{X}(\mathfrak{X}^\top \mathfrak{X})^{-\frac{1}{2}}$.

Outline I: Transformation

- Original model in matrices:

$$\boldsymbol{Y}\boldsymbol{S}(\alpha) = \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k + \boldsymbol{y}_0 \boldsymbol{s}^\top(\alpha) + \Lambda^* \boldsymbol{F}^{*\top} + \varepsilon,$$
$$= \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k + \Lambda \boldsymbol{F}^\top + \varepsilon.$$

- Define the $n \times TK$ matrix $\mathcal{X} := (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K)$, and $\boldsymbol{Q} := \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-\frac{1}{2}}$.
- Transformed model:

$$\tilde{\boldsymbol{Y}}\boldsymbol{S}(\alpha) = \sum_{k=1}^{K} \beta_k \tilde{\boldsymbol{X}}_k + \tilde{\Lambda} \boldsymbol{F}^\top + \tilde{\varepsilon}.$$

Outline I: Transformation

- Original model in matrices:

$$\boldsymbol{Y}\boldsymbol{S}(\alpha) = \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k + \boldsymbol{y}_0 \boldsymbol{s}^\top(\alpha) + \Lambda^* \boldsymbol{F}^{*\top} + \varepsilon,$$
$$= \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k + \Lambda \boldsymbol{F}^\top + \varepsilon.$$

- Define the $n \times TK$ matrix $\mathcal{X} \coloneqq (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K)$, and $\boldsymbol{Q} \coloneqq \mathcal{X}(\mathcal{X}^\top \mathcal{X})^{-\frac{1}{2}}$.
- Transformed model:

$$\tilde{\boldsymbol{Y}}\boldsymbol{S}(\alpha) = \sum_{k=1}^{K} \beta_k \tilde{\boldsymbol{X}}_k + \tilde{\Lambda} \boldsymbol{F}^\top + \tilde{\varepsilon}.$$

- Dimension reduction: $n \times T \to TK \times T$.

Outline II: Objective Function

- Concentrated objective function:

$$\mathcal{Q}(\boldsymbol{\theta}) := \frac{1}{nT} \sum_{r=R+1}^{T} \mu_r \left(\left(\tilde{\boldsymbol{Y}} \boldsymbol{S}(\alpha) - \sum_{k=1}^{K} \beta_k \tilde{\boldsymbol{X}}_k \right)^{\top} \left(\tilde{\boldsymbol{Y}} \boldsymbol{S}(\alpha) - \sum_{k=1}^{K} \beta_k \tilde{\boldsymbol{X}}_k \right) \right).$$

Outline II: Objective Function

- Concentrated objective function:

$$\mathcal{Q}(\boldsymbol{\theta}) := \frac{1}{nT} \sum_{r=R+1}^{T} \mu_r \left(\left(\tilde{\boldsymbol{Y}} \boldsymbol{S}(\alpha) - \sum_{k=1}^{K} \beta_k \tilde{\boldsymbol{X}}_k \right)^{\top} \left(\tilde{\boldsymbol{Y}} \boldsymbol{S}(\alpha) - \sum_{k=1}^{K} \beta_k \tilde{\boldsymbol{X}}_k \right) \right).$$

- Estimator:

$$\hat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta} \in \Theta} \mathcal{Q}(\boldsymbol{\theta}).$$

Assumption MD

Assumption MD (Model)

(i) The parameter vector $\boldsymbol{\theta}^0$ lies in the interior of Θ, where Θ is a compact subset of $\mathbb{R}^{K+1}$ in which $|\alpha| < 1$.

(ii) The elements of the matrices $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K$, Λ^0 and $\boldsymbol{F}^0$ have uniformly bounded fourth moments.

Assumption ER (Error)

(i) $\mathbb{E}[\varepsilon_{it}|\mathcal{C}] = 0$ for $i = 1, \ldots, n$, $t = 1, \ldots, T$.

(ii) Let $\sigma^2_{ij,t\tau} = \mathbb{E}[\varepsilon_{it}\varepsilon_{j\tau}|\mathcal{C}]$. Then $|\sigma^2_{ij,t\tau}| < C$ uniformly for all i, j, t, τ, and the error term is weakly conditionally cross-sectionally and serially dependent, that is, $\sum_{i \neq j} |\sigma^2_{ij,t\tau}| \leq C$ uniformly for all j, t, τ, and $\sum_{t \neq \tau} |\sigma^2_{ij,t\tau}| \leq C$ uniformly for all i, j, τ.

Assumption CS

Assumption CS (Consistency)

(i) $R \geq R^0 := \text{rank}(\tilde{\Lambda}^0 \boldsymbol{F}^{0\top})$.

(ii) $\min_{\boldsymbol{\delta} \in \mathbb{R}^{K+1}:||\boldsymbol{\delta}||_2=1} \sum_{r=R+R^0+1}^{T} \mu_r \left(\frac{1}{nT} (\boldsymbol{\delta} \cdot \tilde{\boldsymbol{Z}})^\top (\boldsymbol{\delta} \cdot \tilde{\boldsymbol{Z}}) \right) \geq b > 0$.

Consistency

Proposition 1 (Consistency)

Under Assumptions MD, ER and CS,

$$||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0||_2 = \mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right).$$

Assumption AE

Assumption AE (Asymptotic Expansion)

(i) $R = R^0$.

(ii) $\frac{1}{n}\tilde{\Lambda}^{0\top}\tilde{\Lambda}^{0} \xrightarrow{p} \Sigma_{\tilde{\Lambda}^0} > 0$ as $n \to \infty$, with $\mu_{R^0}(\Sigma_{\tilde{\Lambda}^0}) > 0$ and $\mu_1(\Sigma_{\tilde{\Lambda}^0}) < \infty$.

(iii) $\frac{1}{T}\boldsymbol{F}^{0\top}\boldsymbol{F}^{0} = \Sigma_{F^0} > 0$, with $\mu_{R^0}(\Sigma_{F^0}) > 0$ and $\mu_1(\Sigma_{F^0}) < \infty$.

Assumption CLT

Assumption CLT (Central Limit Theorem)

$$\frac{1}{\sqrt{nT}}\tilde{\boldsymbol{\mathcal{Z}}}^\top(\boldsymbol{M}_{\boldsymbol{F}^0} \otimes \boldsymbol{M}_{\tilde{\Lambda}^0})\text{vec}(\tilde{\varepsilon}) \xrightarrow{d} \mathcal{N}(0, \Omega),$$

where,

$$\Omega := \frac{1}{nT}\tilde{\boldsymbol{\mathcal{Z}}}^\top(\boldsymbol{M}_{\boldsymbol{F}^0} \otimes \boldsymbol{M}_{\tilde{\Lambda}^0})\Sigma(\boldsymbol{M}_{\boldsymbol{F}^0} \otimes \boldsymbol{M}_{\tilde{\Lambda}^0})\tilde{\boldsymbol{\mathcal{Z}}}.$$

Asymptotic Distribution

Theorem 1 (Asymptotic Distribution)

Under Assumptions MD, CS, AE and CLT, with T fixed and $n \to \infty$,

$$\sqrt{nT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{D}^{-1}\Omega\boldsymbol{D}^{-1}),$$

where,

$$\boldsymbol{D} \coloneqq \frac{1}{nT}\tilde{\boldsymbol{\mathcal{Z}}}^\top(\boldsymbol{M}_{F^0} \otimes \boldsymbol{M}_{\tilde{\Lambda}^0})\tilde{\boldsymbol{\mathcal{Z}}}.$$

Incidental Parameters I

Under some additional assumptions, as $T/n \to c$ with $c \in [0, K^{-1}]$,

$$\sqrt{nT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) + \Delta^{-1}(\boldsymbol{\psi}^{(0)} + \boldsymbol{\psi}^{(1)} + \boldsymbol{\psi}^{(2)} + \boldsymbol{\psi}^{(3)}) \xrightarrow{d} \mathcal{N}(0, \Delta^{-1}\tilde{\Omega}\Delta^{-1}).$$

- $\boldsymbol{\psi}^{(0)}$ and $\boldsymbol{\psi}^{(1)}$ arise due to the presence of the dynamic regressor
- $\boldsymbol{\psi}^{(2)}$ and $\boldsymbol{\psi}^{(3)}$ arise due to cross-sectional dependence/heteroskedasticity and temporal dependence/heteroskedasticity.

Incidental Parameters II

With a few additional assumptions:

	$\psi^{(0)}$	$\psi^{(1)}$	$\psi^{(2)}$	$\psi^{(3)}$
Original	$\mathcal{O}_p\left(\sqrt{\frac{n}{T}}\right)$	$\mathcal{O}_p\left(\sqrt{\frac{n}{T}}\right)$	$\mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right)$	$\mathcal{O}_p\left(\sqrt{\frac{n}{T}}\right)$
Transformed	$\mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right)$	$\mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right)$	$\mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right)$	$\mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right)$.

- **The incidental problem in the cross-section does not disappear entirely, it instead shifts into the time dimension.**

Simulations I

Design:

- True number of factors $R^0 = 2$, factors in estimation $R = 2$.

Simulations I

Design:

- True number of factors $R^0 = 2$, factors in estimation $R = 2$.
- Factors f_{tr} and loadings λ_{ir} drawn from $\mathcal{N}(0, 1)$.

Simulations I

Design:

- True number of factors $R^0 = 2$, factors in estimation $R = 2$.
- Factors f_{tr} and loadings λ_{ir} drawn from $\mathcal{N}(0,1)$.
- Two covariates & a lagged outcome:
  - $\boldsymbol{X}_1 = \Lambda \boldsymbol{F}^\top + \boldsymbol{e}$, where $e_{it} \sim \mathcal{N}(0,1)$.
  - $\boldsymbol{X}_2$ has entries $X_{2,it} \sim \mathcal{N}(0,1)$.

Simulations I

Design:

- True number of factors $R^0 = 2$, factors in estimation $R = 2$.
- Factors f_{tr} and loadings λ_{ir} drawn from $\mathcal{N}(0,1)$.
- Two covariates & a lagged outcome:
    - $\boldsymbol{X}_1 = \Lambda \boldsymbol{F}^\top + \boldsymbol{e}$, where $e_{it} \sim \mathcal{N}(0,1)$.
    - $\boldsymbol{X}_2$ has entries $X_{2,it} \sim \mathcal{N}(0,1)$.
- The entries of the error ε are generated as $\Sigma_n^{\frac{1}{2}} \boldsymbol{U} \Sigma_T^{\frac{1}{2}}$, where
    - $u_{it} \sim \mathcal{N}(0,1)$.
    - Σ_n and Σ_T are diagonal matrices with elements between 0.5 and 2.5.

Simulations I

Design:

- True number of factors $R^0 = 2$, factors in estimation $R = 2$.
- Factors f_{tr} and loadings λ_{ir} drawn from $\mathcal{N}(0,1)$.
- Two covariates & a lagged outcome:
  - $\boldsymbol{X}_1 = \Lambda \boldsymbol{F}^\top + \boldsymbol{e}$, where $e_{it} \sim \mathcal{N}(0,1)$.
  - $\boldsymbol{X}_2$ has entries $X_{2,it} \sim \mathcal{N}(0,1)$.
- The entries of the error ε are generated as $\Sigma_n^{\frac{1}{2}} \boldsymbol{U} \Sigma_T^{\frac{1}{2}}$, where
  - $u_{it} \sim \mathcal{N}(0,1)$.
  - Σ_n and Σ_T are diagonal matrices with elements between 0.5 and 2.5.
- Compare the naive LS estimator (Naive), the LS-IFE estimator applied to the original model (IFE), and the LS-IFE estimator applied to the transformed model (Q-IFE).

Simulations II

Coverage α:

Table 1a: Coverage α %

	Naive			IFE			Q-IFE		
$n \setminus T$	6	9	12	6	9	12	6	9	12
30	85.35	85.30	86.65	60.32	83.73	88.69	59.75	78.99	76.18
60	85.47	86.74	87.31	53.57	79.59	86.39	82.18	84.03	86.08
90	86.47	87.14	87.61	30.72	76.51	83.49	83.72	87.77	86.74
150	86.99	86.26	88.01	22.63	71.12	79.61	84.75	91.32	89.04
300	84.46	87.16	87.70	27.85	62.05	72.27	92.34	92.39	92.11

Simulations III

Coverage β_1:

Table 2a: Coverage β_1 %

	Naive			IFE			Q-IFE		
$n \setminus T$	6	9	12	6	9	12	6	9	12
30	00.84	00.02	00.00	53.68	74.19	80.50	67.00	76.62	78.19
60	00.13	00.00	00.00	62.17	80.23	82.98	81.02	89.55	91.50
90	00.07	00.00	00.00	35.07	81.29	84.92	84.62	91.37	93.03
150	00.03	00.00	00.00	33.31	81.52	87.90	87.86	93.05	93.80
300	00.00	00.00	00.00	44.95	73.44	88.59	90.95	93.42	94.06

Simulations IV

Coverage β_2:

Table 3a: Coverage β_2 %

	Naive			IFE			Q-IFE		
$n \setminus T$	6	9	12	6	9	12	6	9	12
30	86.54	84.87	85.91	84.78	93.09	93.49	88.19	92.71	93,31
60	84.79	86.30	86.42	81.87	92.86	93.62	90.95	93.61	94.16
90	85.72	86.17	87.23	66.25	93.29	94.16	92.85	93.25	94.01
150	88.13	86.56	87.71	53.81	93.17	93.92	92.26	93.36	93.89
300	83.71	87.26	87.41	74.34	93.56	93.55	92.39	93.47	94.02

Conclusion

- More results in the paper:
    - Alternative treatments of the initial condition.
    - Links to classic literature.
    - Low rank regressors.
    - Modified eigenvalue ratio test.

Conclusion

- More results in the paper:
    - Alternative treatments of the initial condition.
    - Links to classic literature.
    - Low rank regressors.
    - Modified eigenvalue ratio test.
- Various extensions.

Extra: LS IFE Estimator

- The LS estimator is obtained as the set of parameter values $(\boldsymbol{\theta}, \Lambda, \boldsymbol{F})$ which minimise the sum of squared residuals.

Extra: LS IFE Estimator

- The LS estimator is obtained as the set of parameter values $(\boldsymbol{\theta}, \Lambda, \boldsymbol{F})$ which minimise the sum of squared residuals.
- It is useful to consider the underlying mechanics:

Extra: LS IFE Estimator

- The LS estimator is obtained as the set of parameter values $(\boldsymbol{\theta}, \Lambda, \boldsymbol{F})$ which minimise the sum of squared residuals.
- It is useful to consider the underlying mechanics:
    1. Given $(\Lambda, \boldsymbol{F})$, estimate $\boldsymbol{\theta}$ by a linear regression.
    2. Given $\boldsymbol{\theta}$, estimate $(\Lambda, \boldsymbol{F})$ by principal components.

Extra: LS IFE Estimator

- The LS estimator is obtained as the set of parameter values $(\boldsymbol{\theta}, \Lambda, \boldsymbol{F})$ which minimise the sum of squared residuals.
- It is useful to consider the underlying mechanics:
  1. Given $(\Lambda, \boldsymbol{F})$, estimate $\boldsymbol{\theta}$ by a linear regression.
  2. Given $\boldsymbol{\theta}$, estimate $(\Lambda, \boldsymbol{F})$ by principal components.
- Problem: estimating n-dimensional factor loadings Λ with T fixed.

Extra: Principal Components

- Let $\dot{\boldsymbol{Y}} := \boldsymbol{Y}\boldsymbol{S}(\alpha) - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k = \Lambda \boldsymbol{F}^\top + \boldsymbol{\varepsilon}$.

Extra: Principal Components

- Let $\dot{\boldsymbol{Y}} := \boldsymbol{Y}\boldsymbol{S}(\alpha) - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k = \Lambda \boldsymbol{F}^\top + \varepsilon$.
- As $n \to \infty$,

$$\frac{1}{nT}\dot{\boldsymbol{Y}}\dot{\boldsymbol{Y}}^\top \check{\Lambda} = \frac{1}{nT}\check{\Lambda}\check{\boldsymbol{F}}^\top \check{\boldsymbol{F}} + \frac{1}{nT}\varepsilon\varepsilon^\top \check{\Lambda} + \boldsymbol{o}_p(1).$$

Extra: Principal Components

- Let $\dot{\boldsymbol{Y}} := \boldsymbol{Y}\boldsymbol{S}(\alpha) - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k = \Lambda \boldsymbol{F}^\top + \varepsilon$.
- As $n \to \infty$,

$$\frac{1}{nT}\dot{\boldsymbol{Y}}\dot{\boldsymbol{Y}}^\top \check{\Lambda} = \frac{1}{nT}\check{\Lambda}\check{\boldsymbol{F}}^\top \check{\boldsymbol{F}} + {\color{red}\frac{1}{nT}\varepsilon\varepsilon^\top \check{\Lambda}} + \boldsymbol{o}_p(1).$$

- Transformed analogue:

$$\frac{1}{nT}\dot{\tilde{\boldsymbol{Y}}}\dot{\tilde{\boldsymbol{Y}}}^\top \check{\tilde{\Lambda}} = \frac{1}{nT}\check{\tilde{\Lambda}}\check{\boldsymbol{F}}^\top \check{\boldsymbol{F}} + \frac{1}{nT}\tilde{\varepsilon}\tilde{\varepsilon}^\top \check{\tilde{\Lambda}} + \boldsymbol{o}_p(1).$$

Extra: Principal Components

- Let $\dot{\boldsymbol{Y}} := \boldsymbol{Y}\boldsymbol{S}(\alpha) - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k = \Lambda \boldsymbol{F}^\top + \varepsilon$.
- As $n \to \infty$,

$$\frac{1}{nT}\dot{\boldsymbol{Y}}\dot{\boldsymbol{Y}}^\top \check{\Lambda} = \frac{1}{nT}\check{\Lambda}\check{\boldsymbol{F}}^\top \check{\boldsymbol{F}} + \frac{1}{nT}\varepsilon\varepsilon^\top \check{\Lambda} + \boldsymbol{o}_p(1).$$

- Transformed analogue:

$$\frac{1}{nT}\dot{\tilde{\boldsymbol{Y}}}\dot{\tilde{\boldsymbol{Y}}}^\top \check{\tilde{\Lambda}} = \frac{1}{nT}\check{\tilde{\Lambda}}\check{\boldsymbol{F}}^\top \check{\boldsymbol{F}} + \frac{1}{nT}\tilde{\varepsilon}\tilde{\varepsilon}^\top \check{\tilde{\Lambda}} + \boldsymbol{o}_p(1).$$

- Now $\frac{1}{nT}\tilde{\varepsilon}\tilde{\varepsilon}^\top = \frac{1}{nT}\varepsilon \boldsymbol{P}_{\mathcal{X}} \varepsilon^\top = \boldsymbol{o}_p(1)$ as $n \to \infty$.

Extra: Principal Components

- Let $\dot{\boldsymbol{Y}} := \boldsymbol{Y}\boldsymbol{S}(\alpha) - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k = \Lambda \boldsymbol{F}^\top + \varepsilon$.
- As $n \to \infty$,

$$\frac{1}{nT}\dot{\boldsymbol{Y}}\dot{\boldsymbol{Y}}^\top \check{\Lambda} = \frac{1}{nT}\check{\Lambda}\check{\boldsymbol{F}}^\top \check{\boldsymbol{F}} + \frac{1}{nT}\varepsilon\varepsilon^\top \check{\Lambda} + \boldsymbol{o}_p(1).$$

- Transformed analogue:

$$\frac{1}{nT}\dot{\tilde{\boldsymbol{Y}}}\dot{\tilde{\boldsymbol{Y}}}^\top \check{\tilde{\Lambda}} = \frac{1}{nT}\check{\tilde{\Lambda}}\check{\boldsymbol{F}}^\top \check{\boldsymbol{F}} + \frac{1}{nT}\tilde{\varepsilon}\tilde{\varepsilon}^\top \check{\tilde{\Lambda}} + \boldsymbol{o}_p(1).$$

- Now $\frac{1}{nT}\tilde{\varepsilon}\tilde{\varepsilon}^\top = \frac{1}{nT}\varepsilon \boldsymbol{P}_{\mathcal{X}} \varepsilon^\top = \boldsymbol{o}_p(1)$ as $n \to \infty$.
- Potentially generous dependence and/or heteroskedasticity in the error.