Panel Data Models with Interactive Fixed Effects and Relatively Small T

Ayden Higgins

October 2025

Panel Models with Interactive Effects

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

$$\boldsymbol{y}_t = \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t.$$

- Parameter of interest $\boldsymbol{\beta}$.

Panel Models with Interactive Effects

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

$$\boldsymbol{y}_t = \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t.$$

- Parameter of interest $\boldsymbol{\beta}$.
- Macro settings: $\boldsymbol{f}_t$ common shocks with heterogeneous effect $\boldsymbol{\Lambda}$.
- Micro settings: time-varying unobserved heterogeneity, e.g. individual, time or group effects:

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

The Least Squares Estimator

- Treating both the factor and loadings as fixed effects allows for correlation between the factors, the loadings and the covariates.
- Bai (2009) studies the least squares (LS) estimator (also called the principal components estimator) and finds that it is consistent and asymptotically normal as $n, T \to \infty$, but is asymptotically biased.
- The latter arises as a consequence of the incidental parameter problem.
- This problem is often felt most acutely in short panel and with T fixed the LS estimator is inconsistent..

Short T Estimators

- Implicitly of explicitly rely on the existence of correlation between the factor term and the covariates. Two distinct strands:
    - Common Correlated Effects: Pesaran (2006), Bai and Li (2014), Everaert and Groote (2016), Westerlund and Urbain (2015).
    - ‘Correlated Effects’: Holtz-Eakin et al., (1988), Ahn et al., (2001, 2013), Robertson and Sarafdis (2015), Juodis and Sarafdis (2022a).

Short T Estimators

- Implicitly of explicitly rely on the existence of correlation between the factor term and the covariates. Two distinct strands:
  - Common Correlated Effects: Pesaran (2006), Bai and Li (2014), Everaert and Groote (2016), Westerlund and Urbain (2015).
  - 'Correlated Effects': Holtz-Eakin et al., (1988), Ahn et al., (2001, 2013), Robertson and Sarafdis (2015), Juodis and Sarafdis (2022a).
- This paper: introduces a *transformed* least squares (TLS) estimator that can be used to estimate panel model with interactive effects when T is small relative to n.
  - (Specific): Consistent and asymptotically unbiased when T is small relative to n.
  - (General): Insight into the relationship between the LS and GMM approaches to panel data models with interactive fixed effects.
  - (Miscellanea): Dynamic models, Inference, ER test.

Outline I

- Original model in matrices:

$$\boldsymbol{Y} = \boldsymbol{X} \cdot \boldsymbol{\beta} + \boldsymbol{\Lambda} \boldsymbol{F}^\top + \boldsymbol{\varepsilon}.$$

Outline I

- Original model in matrices:

$$\boldsymbol{Y} = \boldsymbol{X} \cdot \boldsymbol{\beta} + \boldsymbol{\Lambda} \boldsymbol{F}^\top + \boldsymbol{\varepsilon}.$$

- Define the $n \times TK$ matrix $\mathfrak{X} \coloneqq (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K)$ (assume $\mathfrak{X}$ has full rank).

Outline I

- Original model in matrices:

$$\boldsymbol{Y} = \boldsymbol{X} \cdot \boldsymbol{\beta} + \boldsymbol{\Lambda} \boldsymbol{F}^\top + \boldsymbol{\varepsilon}.$$

- Define the $n \times TK$ matrix $\mathfrak{X} \coloneqq (\boldsymbol{X}_1, \dots, \boldsymbol{X}_K)$ (assume $\mathfrak{X}$ has full rank).
- Take a singular value decomposition $\mathfrak{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$ and define the transformation matrix $\boldsymbol{Q}_{\mathfrak{X}} \coloneqq \boldsymbol{U}\boldsymbol{V}^\top$

- Original model in matrices:

$$\boldsymbol{Y} = \boldsymbol{X} \cdot \boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{F}^\top + \boldsymbol{\varepsilon}.$$

- Define the $n \times TK$ matrix $\mathfrak{X} := (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K)$ (assume $\mathfrak{X}$ has full rank).
- Take a singular value decomposition $\mathfrak{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$ and define the transformation matrix $\boldsymbol{Q}_{\mathfrak{X}} := \boldsymbol{U}\boldsymbol{V}^\top$
- Premultiply by $\boldsymbol{Q}_{\mathfrak{X}}^\top$ to give

$$\boldsymbol{Q}_{\mathfrak{X}}^\top \boldsymbol{Y} = \tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}} \cdot \boldsymbol{\beta} + \tilde{\boldsymbol{\Lambda}}\boldsymbol{F}^\top + \tilde{\boldsymbol{\varepsilon}}.$$

- If $TK < n$, dimension reduction: $n \times T \to TK \times T$.

- The transformed model can be estimated by minimising the LS objective function:

$$\mathcal{Q}(\boldsymbol{\beta}, \tilde{\boldsymbol{\Lambda}}, \boldsymbol{F}) \coloneqq \frac{1}{nT}\text{tr}\left(\left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \boldsymbol{\beta} - \tilde{\boldsymbol{\Lambda}}\boldsymbol{F}^\top\right)^\top \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \boldsymbol{\beta} - \tilde{\boldsymbol{\Lambda}}\boldsymbol{F}^\top\right)\right).$$

- The factors and the transformed loadings can be profiled out

$$\mathcal{Q}(\boldsymbol{\beta}) \coloneqq \frac{1}{nT}\sum_{t=R+1}^{T} \mu_t\left(\left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \boldsymbol{\beta}\right)^\top \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \boldsymbol{\beta}\right)\right).$$

- TLS estimator defined as

$$\hat{\boldsymbol{\beta}} \coloneqq \underset{\boldsymbol{\beta} \in \Theta_\beta}{\arg\min}\, \mathcal{Q}(\boldsymbol{\beta}).$$

- The transformed model can be estimated by minimising the LS objective function:

$$\mathcal{Q}(\boldsymbol{\beta}, \tilde{\boldsymbol{\Lambda}}, \boldsymbol{F}) \coloneqq \frac{1}{nT}\mathrm{tr}\left(\left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \boldsymbol{\beta} - \tilde{\boldsymbol{\Lambda}}\boldsymbol{F}^\top\right)^\top \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \boldsymbol{\beta} - \tilde{\boldsymbol{\Lambda}}\boldsymbol{F}^\top\right)\right).$$

- The factors and the transformed loadings can be profiled out

$$\mathcal{Q}(\boldsymbol{\beta}) \coloneqq \frac{1}{nT}\sum_{t=R+1}^{T} \mu_t\left(\left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \boldsymbol{\beta}\right)^\top \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \boldsymbol{\beta}\right)\right).$$

- TLS estimator defined as

$$\hat{\boldsymbol{\beta}} \coloneqq \underset{\boldsymbol{\beta} \in \Theta_\beta}{\arg\min}\, \mathcal{Q}(\boldsymbol{\beta}).$$

- Notice that if $TK \geq n$, $\boldsymbol{Q}_{\mathcal{X}}\boldsymbol{Q}_{\mathcal{X}}^\top = \boldsymbol{I}_n$ and therefore the TLS estimator *is* the LS estimator.

Assumption MD & ER

Let $\mathcal{C}_{nT} := \sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K)$ and $\boldsymbol{\Sigma}_{\mathcal{C}} := \mathbb{E}[\text{vec}(\boldsymbol{\varepsilon})\text{vec}(\boldsymbol{\varepsilon})^\top | \mathcal{C}_{nT}]$.

Assumption MD

(i) The parameter vector $\boldsymbol{\beta}_0$ lies in the interior of Θ_β, where Θ_β is a compact subset of $\mathbb{R}^K$.

(ii) The elements of $\boldsymbol{X}_k$, $\boldsymbol{\Lambda}_0$, and $\boldsymbol{F}_0$ have uniformly bounded fourth moments.

Let $\mathcal{C}_{nT} := \sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K)$ and $\boldsymbol{\Sigma}_{\mathcal{C}} := \mathbb{E}[\text{vec}(\boldsymbol{\varepsilon})\text{vec}(\boldsymbol{\varepsilon})^\top | \mathcal{C}_{nT}]$.

Assumption MD

(i) The parameter vector $\boldsymbol{\beta}_0$ lies in the interior of Θ_β, where Θ_β is a compact subset of $\mathbb{R}^K$.

(ii) The elements of $\boldsymbol{X}_k$, $\boldsymbol{\Lambda}_0$, and $\boldsymbol{F}_0$ have uniformly bounded fourth moments.

Assumption EC

Conditional on $\mathcal{C}_{nT}$, ε_{it} are independent over i, with $\mathbb{E}[\varepsilon_{it} | \mathcal{C}_{nT}] = 0$, and $\mathbb{E}[\varepsilon_{it}^4 | \mathcal{C}_{nT}]$ uniformly bounded. In addition, the eigenvalues of $\boldsymbol{\Sigma}_{\mathcal{C}}$ are uniformly bounded away from zero and from above by a constant.

Let $\tilde{\boldsymbol{X}} \cdot \boldsymbol{\delta} := \sum_{k=1}^{K} \delta_k \tilde{\boldsymbol{X}}_k$ and $T_{\min} := R_e + R_0 + 1$.

Assumption CS

(i) $R_e \geq R_0 := \text{rank}(\boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top)$, where R_e denotes the number of factors used in estimation, and R_e and R_0 are constants that do not depend on sample size.

(ii) $\min_{\boldsymbol{\delta} \in \mathbb{R}^K : \|\boldsymbol{\delta}\|_2 = 1} \sum_{t=T_{\min}}^{T} \mu_t((nT)^{-1}(\tilde{\boldsymbol{X}} \cdot \boldsymbol{\delta})^\top (\tilde{\boldsymbol{X}} \cdot \boldsymbol{\delta})) \geq b > 0$ w.p.a.1 as $n \to \infty$, with $T \geq T_{\min}$ fixed or $T \to \infty$.

Let $\tilde{\boldsymbol{X}} \cdot \boldsymbol{\delta} := \sum_{k=1}^{K} \delta_k \tilde{\boldsymbol{X}}_k$ and $T_{\min} := R_e + R_0 + 1$.

Assumption CS

(i) $R_e \geq R_0 := \text{rank}(\boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top)$, where R_e denotes the number of factors used in estimation, and R_e and R_0 are constants that do not depend on sample size.

(ii) $\min_{\boldsymbol{\delta} \in \mathbb{R}^K : \|\boldsymbol{\delta}\|_2 = 1} \sum_{t=T_{\min}}^{T} \mu_t((nT)^{-1}(\tilde{\boldsymbol{X}} \cdot \boldsymbol{\delta})^\top (\tilde{\boldsymbol{X}} \cdot \boldsymbol{\delta})) \geq b > 0$ w.p.a.1 as $n \to \infty$, with $T \geq T_{\min}$ fixed or $T \to \infty$.

$$
\begin{aligned}
&\min_{\boldsymbol{\delta} \in \mathbb{R}^K : \|\boldsymbol{\delta}\|_2 = 1} \sum_{t=R_e+R_0+1}^{T} \mu_t \left(\frac{1}{nT} (\boldsymbol{\delta} \cdot \tilde{\boldsymbol{X}})^\top (\boldsymbol{\delta} \cdot \tilde{\boldsymbol{X}}) \right) \\
&= \min_{\tilde{\boldsymbol{\Lambda}} \in \mathbb{R}^{TK \times R_e},\ \boldsymbol{F} \in \mathbb{R}^{T \times R_0}} \mu_{\min} \left(\frac{1}{nT} \tilde{\boldsymbol{\mathcal{X}}}^\top (\boldsymbol{M}_{\boldsymbol{F}} \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}}) \tilde{\boldsymbol{\mathcal{X}}} \right) \\
&\geq \min_{\boldsymbol{\Lambda} \in \mathbb{R}^{n \times R_e},\ \boldsymbol{F} \in \mathbb{R}^{T \times R_0}} \mu_{\min} \left(\frac{1}{nT} \boldsymbol{\mathcal{X}}^\top (\boldsymbol{M}_{\boldsymbol{F}} \otimes \boldsymbol{M}_{\boldsymbol{\Lambda}}) \boldsymbol{\mathcal{X}} \right).
\end{aligned}
$$

Proposition 1 (Consistency)

Under Assumptions MD, EC, and CS, $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ as $n \to \infty$, with $T \geq T_{\min}$ fixed or $T \to \infty$.

- Consistent as $n \to \infty$, regardless of whether T is fixed or $T \to \infty$.
- Allows for heteroskedasticity and serial dependence.
- The number of factors used in estimation is no less than the true number.
- Factors may be strong, weak, or non-existent.

Let $\mathcal{D}_{nT} := \mathcal{C}_{nT} \vee \sigma(\tilde{\boldsymbol{\Lambda}}_0, \boldsymbol{F}_0)$.

Assumption ED

Conditional on $\mathcal{D}_{nT}$, ε_{it} are independent over i, with $\mathbb{E}[\varepsilon_{it}|\mathcal{D}_{nT}] = 0$, $\mathbb{E}[\varepsilon_{it}^2|\mathcal{D}_{nT}] > 0$, and $\sup_{\|\boldsymbol{v}\|_2=1} \mathbb{E}[(\boldsymbol{v}^\top \boldsymbol{\varepsilon}_i)^4|\mathcal{D}_{nT}]$ uniformly bounded for $\mathcal{D}_{nT}$-measurable vectors $\boldsymbol{v}$.

- Strengthens Assumption EC strengthens to restrict dependence between the error and the factor term, and imposes more stringent conditions on serial dependence in the error.
- Closely related to Assumption C(iv) in Bai (2009):

$$\frac{1}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \sum_{t_4=1}^{T} \text{cov}(\varepsilon_{it_1}\varepsilon_{it_2}, \varepsilon_{it_3}\varepsilon_{it_4}) = \mathbb{E}[(\boldsymbol{v}^\top \boldsymbol{\varepsilon}_i)^4] - \mathbb{E}[(\boldsymbol{v}^\top \boldsymbol{\varepsilon}_i)^2]^2,$$

  with $\boldsymbol{v} = \boldsymbol{\iota}_T/\sqrt{T}$.

Assumption AE

(i) $R_e = R_0 = \text{rank}(\tilde{\mathbf{\Lambda}}_0 \boldsymbol{F}_0^\top)$.

(ii) $n^{-1}\tilde{\mathbf{\Lambda}}_0^\top \tilde{\mathbf{\Lambda}}_0 \xrightarrow{p} \mathbf{\Sigma}_{\tilde{\mathbf{\Lambda}}_0}$ as $n \to \infty$, with $T \geq R_0 + 1$ fixed or $T \to \infty$, where the eigenvalues of $\mathbf{\Sigma}_{\tilde{\mathbf{\Lambda}}_0}$ are bounded away from zero and from above by a constant.

(iii) For $T \geq R_0 + 1$ fixed, the eigenvalues of $\boldsymbol{F}_0^\top \boldsymbol{F}_0$ are bounded away from zero and from above by a constant, otherwise $T^{-1}\boldsymbol{F}_0^\top \boldsymbol{F}_0 \xrightarrow{p} \mathbf{\Sigma}_{\boldsymbol{F}_0}$ as $T \to \infty$, where the eigenvalues of $\mathbf{\Sigma}_{\boldsymbol{F}_0}$ are bounded away from zero and from above by a constant.

Assumption AE

(i) $R_e = R_0 = \text{rank}(\tilde{\boldsymbol{\Lambda}}_0 \boldsymbol{F}_0^\top)$.

(ii) $n^{-1}\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0 \xrightarrow{p} \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\Lambda}}_0}$ as $n \to \infty$, with $T \geq R_0 + 1$ fixed or $T \to \infty$, where the eigenvalues of $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\Lambda}}_0}$ are bounded away from zero and from above by a constant.

(iii) For $T \geq R_0 + 1$ fixed, the eigenvalues of $\boldsymbol{F}_0^\top \boldsymbol{F}_0$ are bounded away from zero and from above by a constant, otherwise $T^{-1}\boldsymbol{F}_0^\top \boldsymbol{F}_0 \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{F}_0}$ as $T \to \infty$, where the eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{F}_0}$ are bounded away from zero and from above by a constant.

- Suppose $\boldsymbol{\lambda}_{0,i} \sim \text{iid}(\mathbf{0}, \boldsymbol{\Sigma}_\lambda)$, with $\mathbb{E}[\|\boldsymbol{\lambda}_{0,i}\|_2^4]$ uniformly bounded, and are independent of the covariates. As $n \to \infty$, with T fixed or $T \to \infty$ and $T/n \to \gamma \in [0, \infty)$,

$$\frac{1}{n}\boldsymbol{\Lambda}_0^\top \boldsymbol{\Lambda}_0 \xrightarrow{p} \boldsymbol{\Sigma}_\lambda, \qquad \frac{1}{n}\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0 \xrightarrow{p} \min\{1, \gamma K\} \times \boldsymbol{\Sigma}_\lambda.$$

Proposition 2

Assume $\beta \xrightarrow{p} \boldsymbol{\beta}_0$ as $n \to \infty$, with $T \geq T_{\min}$ fixed or $T \to \infty$. Under Assumptions ME, ED, and AE, as $n \to \infty$, with $T \geq T_{\min}$ fixed or $T \to \infty$,

$$\mathcal{Q}(\boldsymbol{\beta}) = \mathcal{Q}(\boldsymbol{\beta}_0) - \frac{2}{\sqrt{nT}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \boldsymbol{d} + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top \boldsymbol{D}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + r(\boldsymbol{\beta}),$$

where $\boldsymbol{d} := \boldsymbol{c} + \boldsymbol{b}^{(1)} + \boldsymbol{b}^{(2)} + \boldsymbol{b}^{(3)}$ with

$$\begin{aligned}
D_{k_1 k_2} &:= \frac{1}{nT}\mathsf{tr}(\tilde{\boldsymbol{X}}_{k_1} \boldsymbol{M}_{\boldsymbol{F}_0} \tilde{\boldsymbol{X}}_{k_2}^\top \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0}) \\
c_k &:= \frac{1}{\sqrt{nT}}\mathsf{tr}(\tilde{\boldsymbol{X}}_k \boldsymbol{M}_{\boldsymbol{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0}) \\
b_k^{(1)} &:= -\frac{1}{\sqrt{nT}}\mathsf{tr}\left(\boldsymbol{M}_{\boldsymbol{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\boldsymbol{X}}_k \boldsymbol{F}_0 (\boldsymbol{F}_0^\top \boldsymbol{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}}\right) \\
b_k^{(2)} &:= -\frac{1}{\sqrt{nT}}\mathsf{tr}\left(\boldsymbol{M}_{\boldsymbol{F}_0} \tilde{\boldsymbol{X}}_k^\top \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \boldsymbol{F}_0 (\boldsymbol{F}_0^\top \boldsymbol{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\varepsilon}}\right) \\
b_k^{(3)} &:= -\frac{1}{\sqrt{nT}}\mathsf{tr}\left(\boldsymbol{M}_{\boldsymbol{F}_0} \tilde{\boldsymbol{\varepsilon}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0} \tilde{\boldsymbol{\varepsilon}} \boldsymbol{F}_0 (\boldsymbol{F}_0^\top \boldsymbol{F}_0)^{-1} (\tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{\Lambda}}_0)^{-1} \tilde{\boldsymbol{\Lambda}}_0^\top \tilde{\boldsymbol{X}}_k\right).
\end{aligned}$$

Moreover, $r(\boldsymbol{\beta})$ is $o_p((nT)^{-1}(1 + \sqrt{nT}\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2)^2)$.

Let

$$\boldsymbol{V} := \frac{1}{nT}\tilde{\boldsymbol{\mathcal{X}}}^\top(\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0})\tilde{\boldsymbol{\Sigma}}_{\mathcal{D}}(\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0})\tilde{\boldsymbol{\mathcal{X}}},$$

$\boldsymbol{\Sigma}_{\mathcal{D}} := \mathbb{E}[\text{vec}(\boldsymbol{\varepsilon})\text{vec}(\boldsymbol{\varepsilon})^\top|\mathcal{D}_{nT}]$, and $\tilde{\boldsymbol{\Sigma}}_{\mathcal{D}} := (\boldsymbol{I}_T \otimes \boldsymbol{Q}_{\mathcal{X}}^\top)\boldsymbol{\Sigma}_{\mathcal{D}}(\boldsymbol{I}_T \otimes \boldsymbol{Q}_{\mathcal{X}})$.

Assumption AD

(i) The elements of $\boldsymbol{M}_{\boldsymbol{P}_{\mathcal{X}}\boldsymbol{\Lambda}_0}\boldsymbol{X}_k$, $\boldsymbol{\Lambda}_0$, and $\boldsymbol{F}_0$ have uniformly bounded eighth moments.

(ii) There exist nonstochastic matrices $\mathbb{D}$ and $\mathbb{V}$ such that $\boldsymbol{D} \xrightarrow{p} \mathbb{D}$ and $\boldsymbol{V} \xrightarrow{p} \mathbb{V}$ as $n \to \infty$, with $T \geq T_{\min}$ fixed or $T \to \infty$, and the eigenvalues of $\mathbb{D}$ and $\mathbb{V}$ are bounded away from zero and from above by a constant.

Theorem 1 (Asymptotic Distribution)

Assume $\|\boldsymbol{c}\|_2 = \mathcal{O}_p(1)$. Under Assumptions MD, EC, CS, AE, and AD, as $n \to \infty$,

(i) *with $T \geq T_{\min}$ fixed or $T \to \infty$ and $T/n \to 0$,*

$$\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbb{D}^{-1}\mathbb{V}\mathbb{D}^{-1}),$$

(ii) *with $T \to \infty$ and $T/n \to \gamma \in (0, \infty)$,*

$$\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \boldsymbol{D}^{-1}(\boldsymbol{\psi}^{(1)} + \boldsymbol{\psi}^{(2)}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbb{D}^{-1}\mathbb{V}\mathbb{D}^{-1}),$$

where

$$\psi_k^{(1)} := \frac{1}{\sqrt{nT}}\text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{D}}(\boldsymbol{I}_T \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0}\tilde{\boldsymbol{X}}_k\boldsymbol{F}_0(\boldsymbol{F}_0^\top\boldsymbol{F}_0)^{-1}(\tilde{\boldsymbol{\Lambda}}_0^\top\tilde{\boldsymbol{\Lambda}}_0)^{-1}\tilde{\boldsymbol{\Lambda}}_0^\top))$$

$$\psi_k^{(2)} := \frac{1}{\sqrt{nT}}\text{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{D}}(\boldsymbol{F}_0(\boldsymbol{F}_0^\top\boldsymbol{F}_0)^{-1}(\tilde{\boldsymbol{\Lambda}}_0^\top\tilde{\boldsymbol{\Lambda}}_0)^{-1}\tilde{\boldsymbol{\Lambda}}_0^\top\tilde{\boldsymbol{X}}_k\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{I}_{TK})).$$

Asymptotic Distribution II

Theorem 3 of Bai (2009), translated:

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{\mathsf{LS}} - \boldsymbol{\beta}_0) + \boldsymbol{D}_{\mathsf{LS}}^{-1}(\boldsymbol{\psi}_{\mathsf{LS}}^{(1)} + \boldsymbol{\psi}_{\mathsf{LS}}^{(2)}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbb{D}_{\mathsf{LS}}^{-1}\mathbb{V}_{\mathsf{LS}}\mathbb{D}_{\mathsf{LS}}^{-1}),$$

as $n, T \to \infty$, $T/n \to \gamma$ and $\gamma \in (0, \infty)$, and where

$$\psi_{\mathsf{LS},k}^{(1)} := \frac{1}{\sqrt{nT}}\mathsf{tr}(\boldsymbol{\Sigma}_{\mathcal{D}^*}(\boldsymbol{I}_T \otimes \boldsymbol{M}_{\boldsymbol{\Lambda}_0}\boldsymbol{X}_k\boldsymbol{F}_0(\boldsymbol{F}_0^\top\boldsymbol{F}_0)^{-1}(\boldsymbol{\Lambda}_0^\top\boldsymbol{\Lambda}_0)^{-1}\boldsymbol{\Lambda}_0^\top))$$
$$\psi_{\mathsf{LS},k}^{(2)} := \frac{1}{\sqrt{nT}}\mathsf{tr}(\boldsymbol{\Sigma}_{\mathcal{D}^*}(\boldsymbol{F}_0(\boldsymbol{F}_0^\top\boldsymbol{F}_0)^{-1}(\boldsymbol{\Lambda}_0^\top\boldsymbol{\Lambda}_0)^{-1}\boldsymbol{\Lambda}_0^\top\boldsymbol{X}_k\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{I}_n)).$$

Asymptotic Distribution III

- Observe that when $n, T \to \infty$ and $T/n \to \gamma \in [K^{-1}, \infty)$ the LS estimator and the TLS estimator are asymptotically equivalent because with $TK \geq n$, $\boldsymbol{Q}_{\mathcal{X}} \boldsymbol{Q}_{\mathcal{X}}^{\top} = \boldsymbol{I}_n$.

- Observe that when $n, T \to \infty$ and $T/n \to \gamma \in [K^{-1}, \infty)$ the LS estimator and the TLS estimator are asymptotically equivalent because with $TK \geq n$, $\boldsymbol{Q}_{\mathcal{X}} \boldsymbol{Q}_{\mathcal{X}}^\top = \boldsymbol{I}_n$.
- When $T/n \to \gamma \in [0, K^{-1})$

	$\boldsymbol{\psi}_{\bullet}^{(1)}$	$\boldsymbol{\psi}_{\bullet}^{(2)}$
LS Estimator	$\mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right)$	$\mathcal{O}_p\left(\sqrt{\frac{n}{T}}\right)$
TLS Estimator	$\mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right)$	$\mathcal{O}_p\left(\min\left\{\sqrt{\frac{n}{T}}, \sqrt{\frac{T}{n}}\right\}\right)$.

A natural question arises as to the relative efficiency of the LS and TLS estimators. There is no clear ordering, in general. Nonetheless, insight can be gained from considering the case in which the errors are homoskedastic.

Proposition 3

Assume $\mathbf{\Sigma}_{\mathcal{D}} = \mathbf{\Sigma}_{\mathcal{D}^*} = \sigma_0^2 \boldsymbol{I}_{nT}$, *and there exist nonstochastic matrices* $\mathbb{D}$ *and* $\mathbb{D}_{\mathsf{LS}}$ *such that* $\boldsymbol{D} \xrightarrow{p} \mathbb{D}$ *and* $\boldsymbol{D}_{\mathsf{LS}} \xrightarrow{p} \mathbb{D}_{\mathsf{LS}}$ *as* $n, T \to \infty$ *with* $T/n \to \gamma \in (0, \infty)$, *and the eigenvalues of* $\mathbb{D}$ *and* $\mathbb{D}_{\mathsf{LS}}$ *are bounded away from zero and from above by a constant. Moreover, assume*

$$\mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)) = \sigma_0^2 \mathbb{D}^{-1}$$
$$\mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}}_{\mathsf{LS}} - \boldsymbol{\beta}_0)) = \sigma_0^2 \mathbb{D}_{\mathsf{LS}}^{-1},$$

where $\mathsf{avar}(\cdot)$ *denotes asymptotic variance. Then*

$$\mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)) \succeq \mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}}_{\mathsf{LS}} - \boldsymbol{\beta}_0)).$$

Asymptotic Distribution IV

Notice:

$$\boldsymbol{D} - \boldsymbol{D}_{\mathsf{LS}} = \frac{1}{nT}\boldsymbol{\mathcal{X}}^\top(\boldsymbol{M}_{\boldsymbol{F}_0} \otimes (\boldsymbol{P}_{\boldsymbol{\Lambda}_0} - \boldsymbol{P}_{\boldsymbol{P}_{\mathfrak{X}}\boldsymbol{\Lambda}_0}))\boldsymbol{\mathcal{X}}.$$

- Information about the original factor loadings is lost which, ultimately, may result in a larger variance.
- If, for example, $\boldsymbol{\Lambda}_0 = f(\mathfrak{X})$, one may consider a better approximation.
- Include additional external variables in a larger matrix $\boldsymbol{\mathcal{W}}$ to achieve a better approximation of the column space of the factor loadings.
- If $\mathsf{col}((\mathfrak{X}, \boldsymbol{\Lambda}_0)) \subseteq \mathsf{col}(\mathcal{W})$ there is no loss of information in transforming model through $\boldsymbol{Q}_{\mathcal{W}}$.
- There will typically be a cost, since the analogue of the bias $\boldsymbol{\psi}_2$ that appears in Theorem 1 would generally be of order $\min\{n, d\}(nT)^{-\frac{1}{2}}$.

Following on from this discussion, it is natural to compare a generalised least squares interactive fixed effects estimator (GLS) constructed as

$$\hat{\boldsymbol{\beta}}^*_{\mathsf{GLS}} = \left(\boldsymbol{\mathcal{X}}^\top \left((\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{M}_{\boldsymbol{\Lambda}_0})\boldsymbol{\Sigma}_{\mathcal{D}^*}(\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{M}_{\boldsymbol{\Lambda}_0})\right)^+ \boldsymbol{\mathcal{X}}\right)^{-1} \times \boldsymbol{\mathcal{X}}^\top \left((\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{M}_{\boldsymbol{\Lambda}_0})\boldsymbol{\Sigma}_{\mathcal{D}^*}(\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{M}_{\boldsymbol{\Lambda}_0})\right)^+ \mathsf{vec}(\boldsymbol{Y}),$$

and a corresponding generalised *transformed* least squares interactive fixed effects estimator (GTLS)

$$\hat{\boldsymbol{\beta}}^*_{\mathsf{GTLS}} = \left(\tilde{\boldsymbol{\mathcal{X}}}^\top \left((\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0})\tilde{\boldsymbol{\Sigma}}_{\mathcal{D}}(\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0})\right)^+ \tilde{\boldsymbol{\mathcal{X}}}\right)^{-1} \times \tilde{\boldsymbol{\mathcal{X}}}^\top \left((\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0})\tilde{\boldsymbol{\Sigma}}_{\mathcal{D}}(\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0})\right)^+ \mathsf{vec}(\tilde{\boldsymbol{Y}}).$$

Asymptotic Distribution VI

Proposition 4

Assume $\boldsymbol{\Sigma}_{\mathcal{D}} = \boldsymbol{\Sigma}_{\mathcal{D}^*} = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is nonstochastic, and there exist nonstochastic matrices $\mathbb{D}^*$ and $\mathbb{D}^*_{\mathsf{LS}}$, such that $\boldsymbol{D}^* \xrightarrow{p} \mathbb{D}^*$ and $\boldsymbol{D}^*_{\mathsf{LS}} \xrightarrow{p} \mathbb{D}^*_{\mathsf{LS}}$ as $n, T \to \infty$ with $T/n \to \gamma \in (0, \infty)$, and the eigenvalues of $\mathbb{D}^*$ and $\mathbb{D}^*_{\mathsf{LS}}$ are bounded away from zero and from above by a constant. Moreover, assume

$$\mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}}^*_{\mathsf{GTLS}} - \boldsymbol{\beta}_0)) = \mathbb{D}^{*-1}$$
$$\mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}}^*_{\mathsf{GLS}} - \boldsymbol{\beta}_0)) = \mathbb{D}^{*-1}_{\mathsf{LS}}.$$

Then

$$\mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}}^*_{\mathsf{GTLS}} - \boldsymbol{\beta}_0)) \succeq \mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}}^*_{\mathsf{GLS}} - \boldsymbol{\beta}_0)).$$

- According to Theorem 1, with T fixed and $n \to \infty$ the TLS estimator is consistent and asymptotically unbiased.
- Useful to compare with other closely related estimators, in particular the FIVU estimator of Robertson and Sarafidis (2015) (abbreviated to RS), and the quasi-difference estimator of Ahn et al. (2017) (abbreviated to ALS).

RS Estimator I

- Under strict exogeneity

$$\mathbb{E}\left[(\boldsymbol{I}_T \otimes \mathcal{X})^\top \text{vec}(\boldsymbol{\varepsilon})\right] = \mathbb{E}\left[(\boldsymbol{I}_T \otimes \mathcal{X})^\top \text{vec}(\boldsymbol{Y} - \boldsymbol{X} \cdot \boldsymbol{\beta}_0 - \boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top)\right] = \mathbf{0}.$$

- If, in addition, one assumes the data $\{\boldsymbol{X}_i, \boldsymbol{\lambda}_{0,i}, \boldsymbol{\varepsilon}_i\}$ are identically and independently distributed over i, and that the factors are fixed, then

$$\begin{aligned}\mathbb{E}\left[\mathcal{X}^\top \boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top\right] &= \sum_{i=1}^{n} \mathbb{E}\left[\text{vec}(\boldsymbol{X}_i)\boldsymbol{\lambda}_{0,i}^\top\right] \boldsymbol{F}_0^\top \\ &=: n\boldsymbol{\Psi}_0 \boldsymbol{F}_0^\top.\end{aligned}$$

- Th FIVU estimator of Robertson and Sarafidis (2015) is based on the moment condition

$$\mathbb{E}\left[(\boldsymbol{I}_T \otimes \mathcal{X})^\top \text{vec}(\boldsymbol{Y} - \boldsymbol{X} \cdot \boldsymbol{\beta}_0) - n\text{vec}(\boldsymbol{\Psi}_0 \boldsymbol{F}_0^\top)\right] = \mathbf{0}.$$

RS II

- Instead adopt a different perspective:

$$\sum_{i=1}^{n} \text{vec}(\boldsymbol{X}_i)\boldsymbol{\lambda}_{0,i}^{\top} = \mathcal{X}^{\top}\boldsymbol{\Lambda}_0 = (\mathcal{X}^{\top}\mathcal{X})^{\frac{1}{2}}\tilde{\boldsymbol{\Lambda}}_0, \tag{1}$$

  and so consider the alternate moment condition

$$\mathbb{E}\left[(\boldsymbol{I}_T \otimes \mathcal{X})^{\top}\text{vec}(\boldsymbol{Y} - \boldsymbol{X}\cdot\boldsymbol{\beta}_0) - \text{vec}((\mathcal{X}^{\top}\mathcal{X})^{\frac{1}{2}}\tilde{\boldsymbol{\Lambda}}_0\boldsymbol{F}_0^{\top})\right] = \boldsymbol{0}. \tag{M-RS}$$

- M-RS does not identify $\tilde{\boldsymbol{\Lambda}}_0$ nor $\boldsymbol{F}_0$ as $\tilde{\boldsymbol{\Lambda}}_0\boldsymbol{F}_0^{\top} = \tilde{\boldsymbol{\Lambda}}_0\boldsymbol{H}\boldsymbol{H}^{-1}\boldsymbol{F}_0^{\top} = \tilde{\boldsymbol{\Lambda}}_*\boldsymbol{F}_*^{\top}$ for any $R_0 \times R_0$ invertible matrix $\boldsymbol{H}$. The following normalisation is adopted:

$$\boldsymbol{F}_{\text{RS}} = \begin{pmatrix} \boldsymbol{I}_{R_0} \\ \boldsymbol{\Phi}_{\text{RS}} \end{pmatrix}, \qquad \tilde{\boldsymbol{\Lambda}}_{\text{RS}} \text{ is unrestricted.} \tag{R-RS}$$

- The RS estimator is obtained as

$$(\hat{\boldsymbol{\beta}}_{\text{RS}}, \hat{\tilde{\boldsymbol{\Lambda}}}_{\text{RS}}, \hat{\boldsymbol{F}}_{\text{RS}}) := \underset{\boldsymbol{\beta}\in\Theta_{\beta},\ \tilde{\boldsymbol{\Lambda}}\in\Theta_{\tilde{\Lambda}},\ \boldsymbol{F}\in\bar{\Theta}_F}{\arg\min} \mathcal{Q}_{\text{RS}}(\boldsymbol{\beta}, \tilde{\boldsymbol{\Lambda}}, \boldsymbol{F}).$$

- A different moment condition is studied by Ahn et al. (2013) which takes the form

$$\mathbb{E}\left[(\boldsymbol{\mathcal{V}}_0 \otimes \boldsymbol{\mathcal{X}})^\top \text{vec}(\boldsymbol{Y} - \boldsymbol{X} \cdot \boldsymbol{\beta}_0)\right] = \mathbf{0}, \tag{M-ALS}$$

  where the $T \times (T - R_0)$ matrix $\boldsymbol{\mathcal{V}}_0$ forms a basis for the left null space of $\boldsymbol{F}_0$.

- As previously, M-ALS fails to uniquely identify $\boldsymbol{\mathcal{V}}_0$: additional restrictions are adopted. Ahn et al. (2013) consider the following restriction:

$$\boldsymbol{\mathcal{V}} = \begin{pmatrix} \boldsymbol{\Phi}_{\mathsf{ALS}} \\ -\boldsymbol{I}_{T-R_0} \end{pmatrix}. \tag{R-ALS}$$

- The ALS estimator is obtained as

$$(\hat{\boldsymbol{\beta}}_{\mathsf{ALS}}, \hat{\boldsymbol{\mathcal{V}}}_{\mathsf{ALS}}) \coloneqq \underset{\boldsymbol{\beta} \in \in \Theta_{\beta},\ \boldsymbol{\mathcal{V}} \in \bar{\Theta}_{\mathcal{V}}}{\arg\min}\ \mathcal{Q}_{\mathsf{ALS}}(\boldsymbol{\beta}, \boldsymbol{\mathcal{V}}).$$

TLS

- The TLS estimator can be obtained from the moment condition

$$\mathbb{E}\left[(\boldsymbol{\mathcal{V}}_0 \otimes \boldsymbol{Q}_{\mathcal{X}})^\top \mathsf{vec}(\boldsymbol{Y} - \boldsymbol{X} \cdot \boldsymbol{\beta}_0)\right] = \mathbf{0}. \tag{M-TLS}$$

- The TLS estimator tacitly imposes an alternative restriction to R-ALS which takes the form

$$\boldsymbol{\mathcal{V}}^\top \boldsymbol{\mathcal{V}} = \boldsymbol{I}_{T-R_0}. \tag{R-TLS}$$

- Under R-TLS it is possible to profile $\boldsymbol{\mathcal{V}}$ out of the objective function to obtain

$$\begin{aligned}\hat{\boldsymbol{\beta}} &:= \underset{\boldsymbol{\beta} \in \Theta_\beta}{\arg\min} \left(\min_{\boldsymbol{\mathcal{V}} \in \tilde{\Theta}_{\mathcal{V}}} \mathcal{Q}_{\mathsf{TLS}}(\boldsymbol{\beta}, \boldsymbol{\mathcal{V}}) \right) \\ &= \underset{\boldsymbol{\beta} \in \Theta_\beta}{\arg\min} \ \mathcal{Q}(\boldsymbol{\beta}).\end{aligned}$$

Comments

- Suppose $\varepsilon_{it} \sim \text{iid}(0, \sigma_0^2)$ conditional on $\mathcal{D}_{nT}$, then the optimal weighting matrix associated with M-TLS is

$$\boldsymbol{W}^* = \frac{1}{\sigma_0^2}(\boldsymbol{\mathcal{V}}_0^\top \boldsymbol{\mathcal{V}}_0 \otimes \boldsymbol{Q}_{\mathcal{X}}^\top \boldsymbol{Q}_{\mathcal{X}})^{-1} = \frac{1}{\sigma_0^2}\boldsymbol{I}_{(T-R_0)TK}.$$

Comments

- Suppose $\varepsilon_{it} \sim \text{iid}(0, \sigma_0^2)$ conditional on $\mathcal{D}_{nT}$, then the optimal weighting matrix associated with M-TLS is
$$\boldsymbol{W}^* = \frac{1}{\sigma_0^2}(\boldsymbol{\mathcal{V}}_0^\top \boldsymbol{\mathcal{V}}_0 \otimes \boldsymbol{Q}_{\mathcal{X}}^\top \boldsymbol{Q}_{\mathcal{X}})^{-1} = \frac{1}{\sigma_0^2}\boldsymbol{I}_{(T-R_0)TK}.$$

- Transforming the regressors by $\boldsymbol{Q}_{\mathcal{X}}$ tacitly imposes the optimal weighting matrix under homoskedasticity. This is critical to ensuring the estimator remains consistent when $T \to \infty$.

Comments

- Suppose $\varepsilon_{it} \sim \text{iid}(0, \sigma_0^2)$ conditional on $\mathcal{D}_{nT}$, then the optimal weighting matrix associated with M-TLS is

$$\boldsymbol{W}^* = \frac{1}{\sigma_0^2}(\boldsymbol{\mathcal{V}}_0^\top \boldsymbol{\mathcal{V}}_0 \otimes \boldsymbol{Q}_{\mathcal{X}}^\top \boldsymbol{Q}_{\mathcal{X}})^{-1} = \frac{1}{\sigma_0^2}\boldsymbol{I}_{(T-R_0)TK}.$$

- Transforming the regressors by $\boldsymbol{Q}_{\mathcal{X}}$ tacitly imposes the optimal weighting matrix under homoskedasticity. This is critical to ensuring the estimator remains consistent when $T \to \infty$.
- As $n, T \to \infty$ and $T/n \to K^{-1}$, the TLS estimator approaches the LS estimator. Since the LS estimator is known to be consistent when both n and T are large, this closeness is desirable, and is a mirror to the relationship between the within estimator and the optimal GMM estimator discussed in Alvarez and Arellano (2003).

Comments

- Suppose $\varepsilon_{it} \sim \text{iid}(0, \sigma_0^2)$ conditional on $\mathcal{D}_{nT}$, then the optimal weighting matrix associated with M-TLS is

$$\boldsymbol{W}^* = \frac{1}{\sigma_0^2}(\boldsymbol{\mathcal{V}}_0^\top \boldsymbol{\mathcal{V}}_0 \otimes \boldsymbol{Q}_{\mathfrak{X}}^\top \boldsymbol{Q}_{\mathfrak{X}})^{-1} = \frac{1}{\sigma_0^2}\boldsymbol{I}_{(T-R_0)TK}.$$

- Transforming the regressors by $\boldsymbol{Q}_{\mathfrak{X}}$ tacitly imposes the optimal weighting matrix under homoskedasticity. This is critical to ensuring the estimator remains consistent when $T \to \infty$.
- As $n, T \to \infty$ and $T/n \to K^{-1}$, the TLS estimator approaches the LS estimator. Since the LS estimator is known to be consistent when both n and T are large, this closeness is desirable, and is a mirror to the relationship between the within estimator and the optimal GMM estimator discussed in Alvarez and Arellano (2003).
- Comparable one-step ALS and RS estimators are studied. For the RS estimator this amounts to setting $\boldsymbol{W} = (\boldsymbol{I}_T \otimes (\mathfrak{X}^\top \mathfrak{X})^{-1})$, and for the ALS estimator setting $\boldsymbol{W} = (\boldsymbol{I}_{T-R_0} \otimes (\mathfrak{X}^\top \mathfrak{X})^{-1})$.

Proposition 5

Assume it is possible to decompose $\tilde{\mathbf{\Lambda}}_0 \boldsymbol{F}_0^\top = \tilde{\mathbf{\Lambda}}_* \boldsymbol{F}_*^\top$ such that $\boldsymbol{F}_* \in \bar{\Theta}_F$ ... with T fixed and $n \to \infty$,

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{\mathsf{RS}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbb{D}^{-1}\mathbb{V}\mathbb{D}^{-1}).$$

- Proposition 5 establishes that the TLS and one-step RS estimator will share the same asymptotic distribution, despite the RS estimator imposing restrictions of a different nature.
- While R-RS restricts the factors, these will only feature in the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ through the projector $\boldsymbol{M}_{\boldsymbol{F}_*}$.
- This comes with the caveat that it is indeed possible to decompose $\tilde{\mathbf{\Lambda}}_0 \boldsymbol{F}_0^\top = \tilde{\mathbf{\Lambda}}_* \boldsymbol{F}_*^\top$ such that $\boldsymbol{F}_* \in \bar{\Theta}_F$.

Proposition 6

Assume it is possible to decompose $\tilde{\boldsymbol{\Lambda}}_0 \boldsymbol{F}_0^\top = \tilde{\boldsymbol{\Lambda}}_* \boldsymbol{F}_*^\top$ *such that* $\boldsymbol{F}_* \in \bar{\Theta}_F$ *... with* T *fixed and* $n \to \infty$*,*

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{\mathsf{ALS}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbb{D}_*^{-1} \mathbb{V}_* \mathbb{D}_*^{-1}),$$

where

$$\mathbb{D}_* := \operatorname*{plim}_{n\to\infty} \frac{1}{nT} \tilde{\boldsymbol{\mathcal{X}}}^\top (\boldsymbol{\mathcal{V}}_* \boldsymbol{\mathcal{V}}_*^\top \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0}) \tilde{\boldsymbol{\mathcal{X}}}$$
$$\mathbb{V}_* := \operatorname*{plim}_{n\to\infty} \frac{1}{nT} \tilde{\boldsymbol{\mathcal{X}}}^\top (\boldsymbol{\mathcal{V}}_* \boldsymbol{\mathcal{V}}_*^\top \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0}) \boldsymbol{\Sigma}_{\mathcal{D}} (\boldsymbol{\mathcal{V}}_* \boldsymbol{\mathcal{V}}_*^\top \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0}) \tilde{\boldsymbol{\mathcal{X}}}.$$

- The asymptotic distribution of the one-step ALS estimator will generally not coincide with that of the TLS estimator (nor indeed the one-step RS estimator under R-RS), unless $\boldsymbol{\mathcal{V}}_*^\top \boldsymbol{\mathcal{V}}_* = \boldsymbol{I}_{T-R_0}$.
- Notice that Proposition 6 also assumes that it is possible to decompose the factor term in the manner of R-RS.

- Proposition 5 suggest that even under homoskedasticity the asymptotic variance would not collapse to $\mathbb{D}_*^{-1}$ unless $\boldsymbol{\mathcal{V}}_*$ is orthonormal.

Proposition 7

Assume ... and

$$\text{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)) = \mathbb{D}^{-1}\mathbb{V}\mathbb{D}^{-1}$$
$$\text{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}}_{\text{ALS}} - \boldsymbol{\beta}_0)) = \mathbb{D}_*^{-1}\mathbb{V}_*\mathbb{D}_*^{-1}.$$

Then

$$\text{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}}_{\text{ALS}} - \boldsymbol{\beta}_0)) \succeq \text{avar}(\sqrt{nT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)),$$

where avar$(\cdot)$ *denotes asymptotic variance.*

Optimal GMM

- If optimal GMM estimators (denoted by $\hat{\boldsymbol{\beta}}^*_{\mathsf{ALS}}$ and $\hat{\boldsymbol{\beta}}^*_{\mathsf{RS}}$) are considered then the RS and ALS estimators are asymptotically equivalent, indeed

$$
\begin{aligned}
&\operatorname{avar}\left(\sqrt{nT}(\hat{\boldsymbol{\beta}}^*_{\mathsf{ALS}} - \boldsymbol{\beta}_0)\right) \\
&\quad = \operatorname{avar}\left(\sqrt{nT}(\hat{\boldsymbol{\beta}}^*_{\mathsf{RS}} - \boldsymbol{\beta}_0)\right) \\
&\quad = \operatorname{avar}\left(\sqrt{nT}(\hat{\boldsymbol{\beta}}^*_{\mathsf{GTLS}} - \boldsymbol{\beta}_0)\right),
\end{aligned}
$$

  with T fixed and $n \to \infty$.

Two different inferential procedures are proposed: one for large n, fixed T, and one for large n, large T. A single set of conditions is adopted for simplicity. Let $\Gamma_{b_T}(\boldsymbol{A}) := \boldsymbol{A} \odot (\boldsymbol{\Omega} \otimes \boldsymbol{I}_n)$ for an $nT \times nT$ matrix $\boldsymbol{A}$, with $\boldsymbol{\Omega}$ being a $T \times T$ matrix with elements $\omega_{t_1 t_2} = 1\{|t_1 - t_2| < b_T\}$, and b_T is a positive integer-valued sequence.

Assumption IF

(i) The elements of $\boldsymbol{X}_k$ and $\boldsymbol{M}_{\boldsymbol{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0} \boldsymbol{X}_k \boldsymbol{M}_{\boldsymbol{F}_0}$ have uniformly bounded eighth moments.

(ii) $n^{-1} \boldsymbol{\Lambda}_0^\top \boldsymbol{\Lambda}_0 \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}_0}$ as $n \to \infty$, where the eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}_0}$ are bounded from above by a constant.

(iii) Conditional on $\mathcal{D}_{nT}$, ε_{it} are independent over i, with $\mathbb{E}[\varepsilon_{it} | \mathcal{D}_{nT}] = 0$, $\mathbb{E}[\varepsilon_{it}^2 | \mathcal{D}_{nT}] > 0$, and $\sup_{\|\boldsymbol{v}\|_2 = 1} \mathbb{E}[(\boldsymbol{v}^\top \boldsymbol{\varepsilon}_i)^{16} | \mathcal{D}_{nT}]$ uniformly bounded for $\mathcal{D}_{nT}$-measurable vectors $\boldsymbol{v}$.

(iiii) $\|\Gamma_{b_T}(\boldsymbol{\Sigma}_{\mathcal{D}}) - \boldsymbol{\Sigma}_{\mathcal{D}}\|_2 = \mathcal{O}_p(1)$ as $n, T, b_T \to \infty$ with $T/n \to \gamma \in (0, \infty)$ and $b_T^8 / n \to 0$.

Consider a singular value decomposition $(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \hat{\boldsymbol{\beta}}) =: \sum_{t=1}^{T} s_t \boldsymbol{u}_t \boldsymbol{v}_t^\top$ with singular values $s_T \leq \ldots \leq s_1$. Define $\hat{\hat{\boldsymbol{\Lambda}}} := \sqrt{n}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{R_0})$ and $\hat{\boldsymbol{F}} := (s_1 \boldsymbol{v}_1, \ldots, s_{R_0} \boldsymbol{v}_{R_0})/\sqrt{n}$. Let

$$\hat{\boldsymbol{D}} := \frac{1}{nT} \tilde{\boldsymbol{\mathcal{X}}}^\top (\boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{M}_{\hat{\hat{\boldsymbol{\Lambda}}}}) \tilde{\boldsymbol{\mathcal{X}}}$$

$$\hat{\boldsymbol{V}} := \frac{1}{nT} \tilde{\boldsymbol{\mathcal{X}}}^\top (\boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{M}_{\hat{\hat{\boldsymbol{\Lambda}}}}) \hat{\hat{\boldsymbol{\Sigma}}} (\boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{M}_{\hat{\hat{\boldsymbol{\Lambda}}}}) \tilde{\boldsymbol{\mathcal{X}}},$$

where $\hat{\hat{\boldsymbol{\Sigma}}} := (\boldsymbol{I}_T \otimes \boldsymbol{Q}_{\mathcal{X}}^\top) \Gamma_T(\text{vec}(\hat{\boldsymbol{e}})\text{vec}(\hat{\boldsymbol{e}})^\top)(\boldsymbol{I}_T \otimes \boldsymbol{Q}_{\mathcal{X}})$ and $\hat{\boldsymbol{e}} := (\boldsymbol{Y} - \boldsymbol{X} \cdot \hat{\boldsymbol{\beta}}) \boldsymbol{M}_{\hat{\boldsymbol{F}}}$.

Proposition 8

Under Assumptions MD, CS, AE, AD, and IF, as $n \to \infty$ with $T \geq T_{\min}$ fixed,

$$\|\hat{\boldsymbol{D}} - \boldsymbol{D}\|_2 = \mathcal{O}_p(1)$$

$$\|\hat{\boldsymbol{V}} - \boldsymbol{V}\|_2 = \mathcal{O}_p(1).$$

Consider a singular value decomposition $(\boldsymbol{Y} - \boldsymbol{X} \cdot \hat{\boldsymbol{\beta}}) =: \sum_{t=1}^{T} s_t \boldsymbol{u}_t \boldsymbol{v}_t^\top$ with singular values $s_T \leq \ldots \leq s_1$. Then define $\check{\boldsymbol{F}} := (s_1 \boldsymbol{v}_1, \ldots, s_{R_0} \boldsymbol{v}_{R_0})/\sqrt{n}$,

$$\hat{\boldsymbol{D}} := \frac{1}{nT} \tilde{\boldsymbol{\mathcal{X}}}^\top (\boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{M}_{\hat{\tilde{\boldsymbol{\Lambda}}}}) \tilde{\boldsymbol{\mathcal{X}}}$$

$$\hat{\boldsymbol{V}} := \frac{1}{nT} \tilde{\boldsymbol{\mathcal{X}}}^\top (\boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{M}_{\hat{\tilde{\boldsymbol{\Lambda}}}}) \check{\tilde{\boldsymbol{\Sigma}}} (\boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{M}_{\hat{\tilde{\boldsymbol{\Lambda}}}}) \tilde{\boldsymbol{\mathcal{X}}}$$

$$\hat{\psi}_k^{(1)} := \frac{1}{\sqrt{nT}} \mathrm{tr}(\check{\tilde{\boldsymbol{\Sigma}}} (\boldsymbol{I}_T \otimes \boldsymbol{M}_{\hat{\tilde{\boldsymbol{\Lambda}}}} \tilde{\boldsymbol{X}}_k \hat{\boldsymbol{F}} (\hat{\boldsymbol{F}}^\top \hat{\boldsymbol{F}})^{-1} (\hat{\tilde{\boldsymbol{\Lambda}}}^\top \hat{\tilde{\boldsymbol{\Lambda}}})^{-1} \hat{\tilde{\boldsymbol{\Lambda}}}^\top))$$

$$\hat{\psi}_k^{(2)} := \frac{1}{\sqrt{nT}} \mathrm{tr}(\check{\tilde{\boldsymbol{\Sigma}}} (\hat{\boldsymbol{F}} (\hat{\boldsymbol{F}}^\top \hat{\boldsymbol{F}})^{-1} (\hat{\tilde{\boldsymbol{\Lambda}}}^\top \hat{\tilde{\boldsymbol{\Lambda}}})^{-1} \hat{\tilde{\boldsymbol{\Lambda}}}^\top \tilde{\boldsymbol{X}}_k \boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{I}_{TK})),$$

where $\check{\tilde{\boldsymbol{\Sigma}}} := (\boldsymbol{I}_T \otimes \boldsymbol{Q}_{\mathcal{X}}^\top) \Gamma_{b_T} (\mathrm{vec}(\check{\boldsymbol{e}}) \mathrm{vec}(\check{\boldsymbol{e}})^\top) (\boldsymbol{I}_T \otimes \boldsymbol{Q}_{\mathcal{X}})$ and $\check{\boldsymbol{e}} := (\boldsymbol{Y} - \boldsymbol{X} \cdot \hat{\boldsymbol{\beta}}) \boldsymbol{M}_{\check{\boldsymbol{F}}}$.

Proposition 9

Under Assumptions MD, CS, AE, AD, and IF, as $n, T \to \infty$ with $T/n \to \gamma \in (0, \infty)$,

$$\|\hat{\boldsymbol{\psi}}^{(1)} - \boldsymbol{\psi}^{(1)}\|_2 = \mathcal{O}_p(1)$$

$$\|\hat{\boldsymbol{\psi}}^{(2)} - \boldsymbol{\psi}^{(2)}\|_2 = \mathcal{O}_p(1)$$

$$\|\hat{\boldsymbol{D}} - \boldsymbol{D}\|_2 = \mathcal{O}_p(1)$$

$$\|\hat{\boldsymbol{V}} - \boldsymbol{V}\|_2 = \mathcal{O}_p(1).$$

Let ϱ_n be a sequence depending on n (and possibly also T) that tends towards zero. Define

$$\mu_r^* := \mu_r \left(\frac{1}{nT} \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \hat{\boldsymbol{\beta}} \right)^\top \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}} \cdot \hat{\boldsymbol{\beta}} \right) + \varrho_n^2 \boldsymbol{I}_T \right),$$

that is, μ_r^* is the r-th largest eigenvalue of the bracketed matrix on the right. Thereafter let

$$\mathsf{EigR}(r) := \frac{\mu_r^*}{\mu_{r+1}^*} \text{ for } r = 1, \ldots, T-1.$$

Proposition 10

Assume $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 = \mathcal{O}_p(r_{nT})$ and $r_{nT}, \varrho_n \to 0$ with $\varrho_n^{-1} r_{nT} = \mathcal{O}(1)$ as $n \to \infty$, with $T \geq T_{\min}$ fixed or $T \to \infty$. Moreover, assume $R_0 \geq 1$. Under Assumptions MD and AE, as $n \to \infty$, with $T \geq T_{\min}$ fixed or $T \to \infty$,

$$\Pr\left(\max_{1 \leq r \leq T-1} \mathsf{EigR}(r) = R_0 \right) \to 1.$$

Dynamic Model I

- Consider the model

$$\begin{aligned}\boldsymbol{Y} &= \alpha_0 \boldsymbol{Y}_{-1} + \boldsymbol{X} \cdot \boldsymbol{\beta}_0 + \boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top + \boldsymbol{\varepsilon} \\ &=: \boldsymbol{Z} \cdot \boldsymbol{\theta}_0 + \boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top + \boldsymbol{\varepsilon},\end{aligned}$$

  where $\theta_{0,1} := \alpha_0$, $\boldsymbol{Z}_1 := \boldsymbol{Y}_{-1} := (\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{T-1})$, and $\theta_{0,k+1} = \beta_{0,k}$, $\boldsymbol{Z}_{k+1} = \boldsymbol{X}_k$ for $k = 1, \ldots, K$. Let $\mathcal{E}_{nT}$ denote $\sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K, \tilde{\boldsymbol{y}}_0, \tilde{\boldsymbol{\Lambda}}_0, \boldsymbol{F}_0)$.

Dynamic Model I

- Consider the model

$$\begin{aligned}\boldsymbol{Y} &= \alpha_0 \boldsymbol{Y}_{-1} + \boldsymbol{X} \cdot \boldsymbol{\beta}_0 + \boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top + \boldsymbol{\varepsilon} \\ &=: \boldsymbol{Z} \cdot \boldsymbol{\theta}_0 + \boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top + \boldsymbol{\varepsilon},\end{aligned}$$

  where $\theta_{0,1} := \alpha_0$, $\boldsymbol{Z}_1 := \boldsymbol{Y}_{-1} := (\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{T-1})$, and $\theta_{0,k+1} = \beta_{0,k}$, $\boldsymbol{Z}_{k+1} = \boldsymbol{X}_k$ for $k = 1, \ldots, K$. Let $\mathcal{E}_{nT}$ denote $\sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K, \tilde{\boldsymbol{y}}_0, \tilde{\boldsymbol{\Lambda}}_0, \boldsymbol{F}_0)$.
- Assumptions MD, ED, CS, AE, and AD are to be extended to MD*, ED*, CS*, AE*, and AD*.

Dynamic Model I

- Consider the model

$$\begin{aligned}\boldsymbol{Y} &= \alpha_0 \boldsymbol{Y}_{-1} + \boldsymbol{X} \cdot \boldsymbol{\beta}_0 + \boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top + \boldsymbol{\varepsilon} \\ &=: \boldsymbol{Z} \cdot \boldsymbol{\theta}_0 + \boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top + \boldsymbol{\varepsilon},\end{aligned}$$

  where $\theta_{0,1} := \alpha_0$, $\boldsymbol{Z}_1 := \boldsymbol{Y}_{-1} := (\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{T-1})$, and $\theta_{0,k+1} = \beta_{0,k}$, $\boldsymbol{Z}_{k+1} = \boldsymbol{X}_k$ for $k = 1, \ldots, K$. Let $\mathcal{E}_{nT}$ denote $\sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K, \tilde{\boldsymbol{y}}_0, \tilde{\boldsymbol{\Lambda}}_0, \boldsymbol{F}_0)$.
- Assumptions MD, ED, CS, AE, and AD are to be extended to MD*, ED*, CS*, AE*, and AD*.
- Note, Assumption CS*(ii) requires

$$\min_{\boldsymbol{\delta} \in \mathbb{R}^{K+1} : \|\boldsymbol{\delta}\|_2 = 1} \sum_{t=T_{\min}}^{T} \mu_t \left(\frac{1}{nT} (\tilde{\boldsymbol{Z}} \cdot \boldsymbol{\delta})^\top (\tilde{\boldsymbol{Z}} \cdot \boldsymbol{\delta}) \right) \geq b > 0,$$

  w.p.a.1 as $n \to \infty$, with $T \geq T_{\min}$ fixed or $T \to \infty$, and where $\tilde{\boldsymbol{Z}} \cdot \boldsymbol{\delta} := \sum_{k=1}^{K+1} \delta_k \tilde{\boldsymbol{Z}}_k$.

Theorem 2

Assume $\|\boldsymbol{c}_+\|_2 = \mathcal{O}_p(1)$. *Under Assumptions MD*, ED*, CS*, AE*, and AD*, as* $n \to \infty$,

(i) *with* $T \geq T_{\min}$ *fixed or* $T \to \infty$ *and* $T/n \to 0$,

$$\sqrt{nT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbb{D}^{-1}\mathbb{V}\mathbb{D}^{-1}),$$

(ii) *with* $T \to \infty$ *and* $T/n \to \gamma \in (0, \infty)$,

$$\sqrt{nT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{D}_+^{-1}(\boldsymbol{\psi}^{(0)} + \boldsymbol{\psi}^{(1)} + \boldsymbol{\psi}^{(2)}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbb{D}_+^{-1}\mathbb{V}_+\mathbb{D}_+^{-1}),$$

where

$$\boldsymbol{\psi}^{(0)} := \frac{1}{\sqrt{nT}}\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}((\boldsymbol{P}_{\boldsymbol{F}_0}\boldsymbol{G}\boldsymbol{M}_{\boldsymbol{F}_0} + \boldsymbol{G}\boldsymbol{P}_{\boldsymbol{F}_0}) \otimes \boldsymbol{I}_{TK}))\boldsymbol{\pi}_{K+1}$$

$$\psi_k^{(1)} := \frac{1}{\sqrt{nT}}\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}(\boldsymbol{I}_T \otimes \boldsymbol{M}_{\tilde{\boldsymbol{\Lambda}}_0}\tilde{\boldsymbol{H}}_k\boldsymbol{F}_0(\boldsymbol{F}_0^\top\boldsymbol{F}_0)^{-1}(\tilde{\boldsymbol{\Lambda}}_0^\top\tilde{\boldsymbol{\Lambda}}_0)^{-1}\tilde{\boldsymbol{\Lambda}}_0^\top))$$

$$\psi_k^{(2)} := \frac{1}{\sqrt{nT}}\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_{\mathcal{E}}(\boldsymbol{F}_0(\boldsymbol{F}_0^\top\boldsymbol{F}_0)^{-1}(\tilde{\boldsymbol{\Lambda}}_0^\top\tilde{\boldsymbol{\Lambda}}_0)^{-1}\tilde{\boldsymbol{\Lambda}}_0^\top\tilde{\boldsymbol{H}}_k\boldsymbol{M}_{\boldsymbol{F}_0} \otimes \boldsymbol{I}_{TK})).$$

Dynamic Model III

Similar to before:

	$\boldsymbol{\psi}_{\bullet}^{(0)}$	$\boldsymbol{\psi}_{\bullet}^{(1)}$	$\boldsymbol{\psi}_{\bullet}^{(2)}$
LS Estimator	$\mathcal{O}_p\left(\sqrt{\frac{n}{T}}\right)$	$\mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right)$	$\mathcal{O}_p\left(\sqrt{\frac{n}{T}}\right)$
TLS Estimator	$\mathcal{O}_p\left(\min\left\{\sqrt{\frac{n}{T}}, \sqrt{\frac{T}{n}}\right\}\right)$	$\mathcal{O}_p\left(\sqrt{\frac{T}{n}}\right)$	$\mathcal{O}_p\left(\min\left\{\sqrt{\frac{n}{T}}, \sqrt{\frac{T}{n}}\right\}\right)$.

Consider the case in which $\varepsilon_{it} \sim \text{iid}(0, \sigma_0^2)$, and the true factors and loadings take the form of individual effects. In this case $\boldsymbol{\psi}^{(1)} = \boldsymbol{\psi}^{(2)} = \mathbf{0}$ since $\boldsymbol{\Sigma}_{\mathcal{E}} \propto \boldsymbol{I}_{nT}$, leaving the only remaining bias as $\boldsymbol{\psi}^{(0)}$. The expression for $\psi_1^{(0)}$ collapses to

$$\psi_1^{(0)} := \frac{\sigma_0^2}{\sqrt{nT}} \frac{1}{T} \text{tr}(\boldsymbol{P}_{\mathcal{X}}) \text{tr}(\boldsymbol{G} \boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top).$$

A bit of algebra reveals that

$$\psi_1^{(0)} = \min \left\{ \sqrt{\frac{n}{T}}, K\sqrt{\frac{T}{n}} \right\} \frac{\sigma_0^2}{(1-\alpha_0)} \left(1 - \frac{1}{T} \frac{(1-\alpha_0^T)}{1-\alpha_0} \right),$$

which follows because $\text{tr}(\boldsymbol{P}_{\mathcal{X}}) = \min\{n, TK\}$. This again highlights the significance of the transformation $\boldsymbol{Q}_{\mathcal{X}}$. Without this

$$\psi_1^{(0)} = \sqrt{\frac{n}{T}} \frac{\sigma_0^2}{(1-\alpha_0)} \left(1 - \frac{1}{T} \frac{(1-\alpha_0^T)}{1-\alpha_0} \right),$$

which matches (up to scale) the familiar expression derived in Nickell (1981).

Proposition 11

Assume $\boldsymbol{\Sigma}_{\mathcal{E}} = \boldsymbol{\Sigma}_{\mathcal{E}^*} = \sigma_0^2 \boldsymbol{I}_{nT}$, $\|\boldsymbol{y}_0\|_2 = \mathcal{O}_p(\sqrt{n})$, and there exist nonstochastic matrices $\bar{\mathbb{D}}_+$ and $\bar{\mathbb{D}}_{+,\mathsf{LS}}$, such that $\bar{\boldsymbol{D}}_+ \xrightarrow{p} \bar{\mathbb{D}}_+$ and $\bar{\boldsymbol{D}}_{+,\mathsf{LS}} \xrightarrow{p} \bar{\mathbb{D}}_{+,\mathsf{LS}}$ as $n, T \to \infty$ with $T/n \to \gamma \in (0, \infty)$, and the eigenvalues of $\bar{\mathbb{D}}_+$ and $\bar{\mathbb{D}}_{+,\mathsf{LS}}$ are bounded away from zero and from above by a constant. Moreover, assume

$$\mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\theta}}_{\mathsf{TLS}} - \boldsymbol{\theta}_0)) = \sigma_0^2 \bar{\mathbb{D}}_+^{-1}$$
$$\mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\theta}}_{\mathsf{LS}} - \boldsymbol{\theta}_0)) = \sigma_0^2 \bar{\mathbb{D}}_{+,\mathsf{LS}}^{-1},$$

where avar$(\cdot)$ *denotes asymptotic variance. Then*

$$\mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\theta}}_{\mathsf{TLS}} - \boldsymbol{\theta}_0)) \succeq \mathsf{avar}(\sqrt{nT}(\hat{\boldsymbol{\theta}}_{\mathsf{LS}} - \boldsymbol{\theta}_0)).$$

- In this case one can decompose

$$\begin{aligned}\bar{\boldsymbol{D}}_{+,\mathsf{LS}} - \bar{\boldsymbol{D}}_+ = &\frac{1}{nT} \bar{\boldsymbol{\mathcal{H}}}^\top (\boldsymbol{M}_{\boldsymbol{F}_0} \otimes (\boldsymbol{P}_{\boldsymbol{\Lambda}_0} - \boldsymbol{P}_{\boldsymbol{P}_{\mathcal{X}} \boldsymbol{\Lambda}_0})) \bar{\boldsymbol{\mathcal{H}}} \\ &+ \frac{\sigma_0^2}{nT} \mathsf{tr}((\boldsymbol{G}\boldsymbol{G}^\top \otimes \boldsymbol{M}_{\mathcal{X}})) \boldsymbol{\pi}_{K+1} \boldsymbol{\pi}_{K+1}^\top.\end{aligned}$$

Short Panel Exercise I

- This first exercise compares the performance of the TLS estimator with existing alternatives in the context of a short panel.
- Outcomes are generated according to
$$\boldsymbol{Y} = \beta_{0,1}\boldsymbol{X}_1 + \beta_{0,2}\boldsymbol{X}_2 + \boldsymbol{\Lambda}_0\boldsymbol{F}_0^\top + \boldsymbol{\varepsilon},$$
with $\beta_{0,1} = 1$ and $\beta_{0,2} = -1$.
- The covariate $\boldsymbol{X}_1$ is generated with elements drawn from standard normal distributions. The covariate $\boldsymbol{X}_2 = \boldsymbol{\Lambda}_0\boldsymbol{F}_0^\top + \boldsymbol{\epsilon}$, where $R_0 = 2$.
- $\lambda_{0,ir}$, $f_{0,tr}$ and ϵ_{it} are drawn independently from standard normal distributions.
- For the error in the outcome equation, first a variable u_{it} is generated as $u_{it} := u_{it}^* \times \|\boldsymbol{f}_{0,t}\|_2$ where u_{it}^* are independent over i and t and normally distributed with a mean of zero and variance drawn uniformly from the interval $[0.5, 1.5]$. Thereafter, the errors are generated according to $\varepsilon_{it} = \phi\varepsilon_{i,t-1} + u_{it}$ with $\phi = 0.5$ and $\varepsilon_{i0} = 0$.
- The errors exhibit both conditional and unconditional heteroskedasticity, as well as serial correlation.

Short Panel Exercise II

Table 1: Empirical Bias (Empirical Standard Error)

		β_1			β_2		
	$n \setminus T$	6	9	12	6	9	12
LS	100	**0.004**	**-0.002**	**0.017**	**6.417**	**6.185**	**4.395**
		(1.393)	(1.377)	(1.356)	(5.405)	(6.047)	(5.563)
	250	**0.019**	**0.004**	**-0.003**	**9.789**	**8.962**	**5.382**
		(1.358)	(1.334)	(1.338)	(8.091)	(8.757)	(6.915)
	500	**-0.013**	**-0.019**	**-0.023**	**13.782**	**12.008**	**6.614**
		(1.349)	(1.347)	(1.329)	(11.292)	(11.849)	(8.171)
TLS	100	**0.010**	**-0.004**	**0.021**	**0.349**	**0.232**	**0.214**
		(1.474)	(1.396)	(1.369)	(1.933)	(1.545)	(1.450)
	250	**0.014**	**-0.004**	**0.009**	**0.088**	**0.053**	**0.034**
		(1.454)	(1.373)	(1.364)	(1.522)	(1.392)	(1.370)
	500	**-0.011**	**-0.019**	**-0.021**	**0.030**	**0.028**	**0.013**
		(1.455)	(1.385)	(1.353)	(1.477)	(1.395)	(1.355)
ALS	100	**0.031**	**0.003**	**0.019**	**0.356**	**0.173**	**0.191**
		(2.164)	(2.531)	(2.928)	(2.637)	(2.709)	(3.117)
	250	**0.016**	**0.019**	**-0.012**	**0.093**	**0.032**	**0.028**
		(2.126)	(2.548)	(2.870)	(2.224)	(2.528)	(2.868)
	500	**-0.012**	**-0.008**	**-0.014**	**0.025**	**0.041**	**-0.009**
		(2.140)	(2.516)	(2.839)	(2.140)	(2.546)	(2.884)

Short Panel Exercise III

Table 2: Empirical Coverage Probability of a 95% Confidence Interval

		β_1			β_2		
	$n \setminus T$	6	9	12	6	9	12
LS	100	0.887	0.910	0.924	0.221	0.303	0.456
	250	0.901	0.928	0.928	0.122	0.187	0.347
	500	0.903	0.924	0.928	0.067	0.101	0.221
TLS	100	0.942	0.941	0.944	0.892	0.923	0.928
	250	0.946	0.950	0.946	0.940	0.948	0.947
	500	0.948	0.949	0.950	0.945	0.947	0.952
ALS	100	0.940	0.947	0.944	0.910	0.937	0.941
	250	0.948	0.945	0.948	0.943	0.947	0.947
	500	0.947	0.949	0.951	0.946	0.945	0.950

Short Panel Exercise IV

Table 3: Percentage of Estimated R equal to R_0

$n \setminus T$	9	12	15
100	75.84	83.59	87.33
250	89.51	95.63	97.37
500	95.03	98.26	99.38

Large Panel Exercise I

- This second exercise compares the LS and TLS estimators, as well as their bias-corrected counterparts, in a setting where both n and T are large.
- Outcomes are generated according to

$$\boldsymbol{Y} = \alpha_0 \boldsymbol{Y}_{-1} + \beta_{0,1} \boldsymbol{X}_1 + \beta_{0,2} \boldsymbol{X}_2 + \boldsymbol{\Lambda}_0 \boldsymbol{F}_0^\top + \boldsymbol{\varepsilon},$$

  with $\alpha_0 = 0.5$, $\beta_{0,1} = 1$ and $\beta_{0,2} = -1$.
- The regressors, the factors, the loadings and the covariates are all generated in the same manner as in the previous design.
- The error is also generated as previously, but with the autoregressive parameter $\phi = 0$.

Large Panel Exercise II

Table 4: Empirical Bias (Empirical Standard Error)

		α			β_1			β_2		
	$n \setminus T$	*10*	*25*	*50*	*10*	*25*	*50*	*10*	*25*	*50*
LS	*100*	***-0.031***	***-0.006***	***0.003***	***-0.026***	***-0.012***	***-0.007***	***0.125***	***0.050***	***0.021***
		(0.730)	*(0.559)*	*(0.529)*	*(1.143)*	*(1.048)*	*(1.027)*	*(1.152)*	*(1.063)*	*(1.046)*
	250	***-0.054***	***-0.008***	***-0.002***	***-0.018***	***0.000***	***0.001***	***0.106***	***0.039***	***0.018***
		(0.909)	*(0.583)*	*(0.534)*	*(1.124)*	*(1.050)*	*(1.025)*	*(1.117)*	*(1.064)*	*(1.026)*
	500	***-0.057***	***-0.012***	***-0.012***	***-0.063***	***0.002***	***0.006***	***0.111***	***0.030***	***0.023***
		(1.150)	*(0.647)*	*(0.544)*	*(1.138)*	*(1.032)*	*(1.014)*	*(1.132)*	*(1.055)*	*(1.021)*
LS-BC	*100*	***-0.017***	***-0.003***	***0.004***	***-0.008***	***-0.010***	***-0.007***	***0.107***	***0.048***	***0.021***
		(0.594)	*(0.534)*	*(0.524)*	*(1.142)*	*(1.047)*	*(1.027)*	*(1.151)*	*(1.063)*	*(1.046)*
	250	***-0.020***	***-0.005***	***-0.001***	***0.009***	***0.003***	***0.002***	***0.079***	***0.036***	***0.017***
		(0.611)	*(0.527)*	*(0.518)*	*(1.123)*	*(1.050)*	*(1.025)*	*(1.116)*	*(1.064)*	*(1.026)*
	500	***-0.023***	***-0.007***	***-0.009***	***-0.025***	***0.006***	***0.007***	***0.073***	***0.026***	***0.022***
		(0.648)	*(0.533)*	*(0.512)*	*(1.135)*	*(1.032)*	*(1.014)*	*(1.129)*	*(1.055)*	*(1.021)*
TLS	*100*	***-0.021***	***-0.002***	***0.003***	***-0.011***	***-0.011***	***-0.007***	***0.033***	***0.026***	***0.021***
		(0.746)	*(0.603)*	*(0.529)*	*(1.143)*	*(1.047)*	*(1.027)*	*(1.142)*	*(1.061)*	*(1.046)*
	250	***-0.005***	***-0.006***	***-0.002***	***0.015***	***0.002***	***0.001***	***0.019***	***0.013***	***0.006***
		(0.770)	*(0.638)*	*(0.589)*	*(1.124)*	*(1.050)*	*(1.025)*	*(1.110)*	*(1.063)*	*(1.026)*
	500	***-0.010***	***-0.013***	***-0.012***	***-0.015***	***0.006***	***0.007***	***0.023***	***0.007***	***0.012***
		(0.772)	*(0.660)*	*(0.608)*	*(1.136)*	*(1.032)*	*(1.014)*	*(1.125)*	*(1.054)*	*(1.021)*
TLS-BC	*100*	***-0.015***	***0.000***	***0.004***	***-0.006***	***-0.010***	***-0.007***	***0.027***	***0.025***	***0.021***
		(0.739)	*(0.595)*	*(0.524)*	*(1.143)*	*(1.047)*	*(1.027)*	*(1.142)*	*(1.061)*	*(1.046)*
	250	***0.001***	***-0.005***	***-0.001***	***0.019***	***0.003***	***0.002***	***0.016***	***0.013***	***0.006***
		(0.767)	*(0.635)*	*(0.585)*	*(1.124)*	*(1.050)*	*(1.025)*	*(1.110)*	*(1.063)*	*(1.026)*
	500	***-0.007***	***-0.012***	***-0.011***	***-0.012***	***0.007***	***0.007***	***0.020***	***0.007***	***0.012***
		(0.770)	*(0.659)*	*(0.606)*	*(1.136)*	*(1.032)*	*(1.014)*	*(1.125)*	*(1.054)*	*(1.021)*

Table 5: Empirical Coverage Probability of a 95% Confidence Interval

		α			β_1			β_2		
	$n \setminus T$	10	25	50	10	25	50	10	25	50
LS	100	0.842	0.920	0.937	0.921	0.941	0.944	0.914	0.933	0.937
	250	0.744	0.913	0.934	0.924	0.940	0.946	0.926	0.935	0.945
	500	0.630	0.873	0.929	0.920	0.943	0.945	0.922	0.937	0.947
LS-BC	100	0.911	0.932	0.940	0.921	0.941	0.943	0.915	0.933	0.938
	250	0.904	0.937	0.942	0.924	0.940	0.946	0.924	0.935	0.945
	500	0.885	0.938	0.945	0.921	0.942	0.946	0.923	0.938	0.947
TLS (Long)	100	0.916	0.933	0.937	0.923	0.941	0.944	0.921	0.934	0.937
	250	0.919	0.938	0.941	0.924	0.939	0.946	0.928	0.935	0.945
	500	0.926	0.938	0.943	0.921	0.943	0.946	0.926	0.938	0.947
TLS-BC (Long)	100	0.919	0.937	0.940	0.923	0.940	0.943	0.921	0.934	0.938
	250	0.921	0.940	0.943	0.924	0.939	0.946	0.928	0.935	0.945
	500	0.927	0.939	0.944	0.921	0.943	0.946	0.926	0.938	0.947
TLS (Short)	100	0.939	0.937	0.937	0.944	0.945	0.943	0.942	0.942	0.940
	250	0.944	0.946	0.944	0.948	0.947	0.949	0.949	0.943	0.948
	500	0.950	0.946	0.946	0.946	0.952	0.950	0.949	0.946	0.951
TLS-BC (Short)	100	0.939	0.940	0.942	0.944	0.945	0.943	0.942	0.942	0.940
	250	0.943	0.947	0.947	0.948	0.947	0.949	0.949	0.943	0.948
	500	0.951	0.948	0.947	0.946	0.952	0.950	0.949	0.946	0.951

Table 6: Percentage of Estimated R equal to R_0

$n \setminus T$	10	25	50
100	84.59	98.97	99.94
300	94.50	99.96	100.00
500	97.81	99.99	100.00

Closing Remarks

- This paper has introduced a method to estimate linear panel data models with interactive fixed effects designed for situations where T is small relative to n.

Closing Remarks

- This paper has introduced a method to estimate linear panel data models with interactive fixed effects designed for situations where T is small relative to n.
- This approach has been shown to deliver consistent and asymptotically unbiased estimates of regression slope coefficients with T fixed, and also when both n and T are large, provided the ratio $T/n \to 0$.

Closing Remarks

- This paper has introduced a method to estimate linear panel data models with interactive fixed effects designed for situations where T is small relative to n.
- This approach has been shown to deliver consistent and asymptotically unbiased estimates of regression slope coefficients with T fixed, and also when both n and T are large, provided the ratio $T/n \to 0$.
- This stands in contrast to the usual case where the LS estimator generally delivers inconsistent estimates with T fixed, and suffers from asymptotic bias as $n, T \to \infty$.

Closing Remarks

- This paper has introduced a method to estimate linear panel data models with interactive fixed effects designed for situations where T is small relative to n.
- This approach has been shown to deliver consistent and asymptotically unbiased estimates of regression slope coefficients with T fixed, and also when both n and T are large, provided the ratio $T/n \to 0$.
- This stands in contrast to the usual case where the LS estimator generally delivers inconsistent estimates with T fixed, and suffers from asymptotic bias as $n, T \to \infty$.
- Careful study of this estimation approach has also revealed interesting connections between the LS estimator and several method of moments-based approaches, bridging the gap between what are, at present, two quite separate literatures.

Closing Remarks

- This paper has introduced a method to estimate linear panel data models with interactive fixed effects designed for situations where T is small relative to n.
- This approach has been shown to deliver consistent and asymptotically unbiased estimates of regression slope coefficients with T fixed, and also when both n and T are large, provided the ratio $T/n \to 0$.
- This stands in contrast to the usual case where the LS estimator generally delivers inconsistent estimates with T fixed, and suffers from asymptotic bias as $n, T \to \infty$.
- Careful study of this estimation approach has also revealed interesting connections between the LS estimator and several method of moments-based approaches, bridging the gap between what are, at present, two quite separate literatures.
- Ongoing extensions: (i) predetermined regressors, (ii) endogenous regressors.