Online Supplement to An Incidental Parameters Free Inference Approach for Panels with Common Shocks

Juodis, Arturas and Sarafidis, Vasilis

UvA, BI Norwegian Business School

24 December 2020

Online Supplement to An Incidental Parameters Free Inference Approach for Panels with Common Shocks

Artūras Juodis[1], Vasilis Sarafidis[1]

[a] *Amsterdam School of Economics, University of Amsterdam and Tinbergen Institute*
[b] *Department of Economics, BI Norwegian Business School*

Contents

S.1 Fixed T Limit Theory **2**

S.2 Elimination of Common Time Effects **10**

S.3 Identification-Robust Inference **13**

S.4 Identification in the Panel AR(1) Model using Averaged Moment Conditions **14**

S.4.1 Theoretical analysis 14

S.4.2 Simulations 18

Contents

Section S.1 of the present Supplementary Appendix studies the properties of the proposed GMM estimators under fixed T asymptotics. Section S.2 analyses the effect of transforming the model in terms of time-specific cross-sectional averages on the proposed estimating equations. Section S.3 considers identification-robust inference, building upon the idea of Anderson and Rubin (1949) and Stock and Wright (2000). Finally, Section S.4 discusses local and global identification for the panel AR(1) model and reports additional Monte Carlo results for this model.

S.1. Fixed T Limit Theory

In the fixed T literature, it is standard practice to treat $\{f_t\}_{t=1}^T$ as fixed parameters to be estimated, see e.g. Robertson and Sarafidis (2015). However, in order to keep simple the comparison between the fixed T and the large T results of this paper, we deviate from common practice and assume that the factors are random variables, as in Andrews (2005) and Kuersteiner and Prucha (2013; 2020).

Let Ξ be some generic $\mathcal{F}$-measurable integrable function with $\mathrm{E}[|\Xi|] < \infty$. The following assumptions are employed:

Assumption S.1.1 (Fixed T)**.** *The DGP for all i and t satisfies the following restrictions for $r = 4$ and some $\delta > 0$:*

(a) *(Random Sampling) $\boldsymbol{\Upsilon}_i = (\boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i^d, \boldsymbol{q}_i)$ are identically distributed and independent across i, conditional on $\mathcal{F}$.*

(b) *Each time-varying element $p_{i,t}^{(h)}$ in $\boldsymbol{p}_{i,t} = (\boldsymbol{x}_{i,t}', \boldsymbol{z}_{i,t}', \varepsilon_{i,t}, \varepsilon_{i,t}^d, q_{i,t})'$, satisfies $\mathrm{E}_{\mathcal{F}}\left[\left|p_{i,t}^{(h)}\right|^{r+\delta}\right] \leq \Xi$.*

(c) *Each time-invariant element $v_i^{(h)}$ in $\boldsymbol{v}_i = (\lambda_i, \lambda_i^d, q_i)'$ satisfies $\mathrm{E}_{\mathcal{F}}\left[\left|v_i^{(h)}\right|^{r+\delta}\right] \leq \Xi$.*

(d) *$\mathrm{E}_{\mathcal{F}}[\varepsilon_{i,s}|\boldsymbol{z}_{i,t}, \boldsymbol{v}_i] = 0$ and $\mathrm{E}_{\mathcal{F}}[\varepsilon_{i,s}^d|\boldsymbol{v}_i, q_{i,t}] = 0$ for all $s \geq t$.*

Subject to some additional restrictions on the summability of sequence of scalars, the random sampling restriction can be relaxed without affecting our results. For example, the asymptotic properties of the proposed estimator are not affected if $\boldsymbol{v}_i$, and hence λ_i, is treated as a fixed sequence of constants, see Juodis and Sarafidis (2020).

For T fixed, the estimating equations in Eq. (17) of the main paper are unbiased, whether linear or nonlinear. Therefore, in what follows we formulate the problem without the "delete-one" construction of $\widehat{f}_t$, using time-invariant weights, q_i. In practice it is entirely feasible to employ the delete-one construction for any arbitrary value of T. Moreover, we shall derive the asymptotic properties of the Method of Moments estimator based on averaged estimating equations only. This is because for fixed T, the assumptions and proof strategy employed for stacked moment conditions are almost identical. In what follows, in order to simplify some expressions, we shall use the shorthand notation $N_1 = N - 1$ and $T_1 = T - 1$.

Let

$$\overline{\overline{\boldsymbol{m}}}^{(\xi)}(\boldsymbol{\beta}) = \frac{1}{T_1}\sum_{t=1}^{T_1} \overline{\boldsymbol{m}}_t^{(\xi)}(\boldsymbol{\beta}), \tag{S.1}$$

with

$$\overline{\boldsymbol{m}}_t^{(\xi)}(\boldsymbol{\beta}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[w_{j,t+1}^{(\xi)}\boldsymbol{z}_{i,t}\left(y_{i,t} - \boldsymbol{x}_{i,t}'\boldsymbol{\beta}\right) - w_{j,t}^{(\xi)}\boldsymbol{z}_{i,t}\left(y_{i,t+1} - \boldsymbol{x}_{i,t+1}'\boldsymbol{\beta}\right)\right], \tag{S.2}$$

for $t = 1, \ldots, T_1$ and $\xi \in \{L; NL\}$. For the linear estimator we have

$$w_{i,s}^{(L)} = q_i d_{i,s} = q_i\lambda_i^d f_s + q_i\varepsilon_{i,s}^d = c_i f_s + \eta_{i,s}, \tag{S.3}$$

and a similar definition applies to the nonlinear estimator mutatis mutandis. The at-most-quadratic nature of the proposed estimating equations implies that

$$\overline{\overline{\boldsymbol{m}}}^{(\xi)}(\boldsymbol{\beta}) = \overline{\overline{\boldsymbol{m}}}^{(\xi)}(\boldsymbol{\beta}_0) + \left(\overline{\overline{\boldsymbol{\Gamma}}}_T^{(\xi)} + \frac{1}{2}\sum_{k=1}^{K}\overline{\overline{\boldsymbol{H}}}_{T,k}^{(\xi)}(\beta_k - \beta_{0,k})\right)(\boldsymbol{\beta} - \boldsymbol{\beta}_0), \tag{S.4}$$

where $\overline{\overline{\boldsymbol{\Gamma}}}_T^{(\xi)} = \left[\partial\overline{\overline{\boldsymbol{m}}}^{(\xi)}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'\right]_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$, while the $[D \times K]$ matrices $\overline{\overline{\boldsymbol{H}}}_{T,k}^{(\xi)}$ denote the corresponding matrix-valued second derivatives of $\overline{\overline{\boldsymbol{m}}}^{(\xi)}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, where β_k is the k^{th} element of $\boldsymbol{\beta}$.

Assumption S.1.2 (Local Identification)**.** *For each $\xi = \{L; NL\}$ the limiting Jacobian matrix $\boldsymbol{\Gamma}_T^{(\xi)} = \text{plim}_{N\to\infty}\overline{\overline{\boldsymbol{\Gamma}}}_T^{(\xi)}$ is $\mathcal{F}$-measurable and has rank K a.s. for all T.*

Assumption S.1.3 (Global Identification)**.** *The parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^K$ is compact and contains $\boldsymbol{\beta}_0$ in its interior. Let $\boldsymbol{m}_T^{(NL)}(\boldsymbol{\beta}) = \text{plim}_{N\to\infty}\left[\overline{\overline{\boldsymbol{m}}}^{(NL)}(\boldsymbol{\beta})\right]$ for all $\boldsymbol{\beta} \in \boldsymbol{\Theta}$. $\boldsymbol{\beta}_0$ is identified on $\boldsymbol{\Theta}$ such that $\boldsymbol{m}_T^{(NL)}(\boldsymbol{\beta}) = \mathbf{0}_D$ iff $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ a.s.*

For both choices of moment conditions, we define the estimator $\widehat{\boldsymbol{\beta}}^{(\xi)}$ as the solution of the standard GMM minimisation problem:

$$\widehat{\boldsymbol{\beta}}^{(\xi)} = \underset{\boldsymbol{\beta}\in\boldsymbol{\Theta}}{\arg\min}\ \left(\left(\overline{\overline{\boldsymbol{m}}}^{(\xi)}(\boldsymbol{\beta})\right)'\boldsymbol{W}_{N,T}\overline{\overline{\boldsymbol{m}}}^{(\xi)}(\boldsymbol{\beta})\right), \tag{S.5}$$

for some positive definite weighting matrix $\boldsymbol{W}_{N,T}$. Furthermore, $\boldsymbol{W}_{N,T} \xrightarrow{p} \boldsymbol{W}_T$ as $N \to \infty$, where $\boldsymbol{W}_T$ is assumed to be $\mathcal{F}$-measurable and positive definite a.s.

Next, in order to simplify asymptotic expressions define

$$\boldsymbol{\Sigma}_T^{(\xi)} \equiv \underset{N\to\infty}{\text{plim}}\left[\left(\left(\boldsymbol{\Gamma}_T^{(\xi)}\right)'\boldsymbol{W}_{N,T}\boldsymbol{\Gamma}_T^{(\xi)}\right)^{-1}\left(\boldsymbol{\Gamma}_T^{(\xi)}\right)'\boldsymbol{W}_{N,T}\right], \tag{S.6}$$

where again this quantity is assumed to be $\mathcal{F}-$measurable. Let

$$\boldsymbol{\Omega}_T^{(\xi)} = \text{E}_{\mathcal{F}}\left[(T_1)\boldsymbol{\mu}_{i,T}^{(\xi)}\left(\boldsymbol{\mu}_{i,T}^{(\xi)}\right)'\right] \tag{S.7}$$

be a $[D \times D]$ matrix that is full rank a.s. The following theorem summarises the asymptotic properties of $\widehat{\boldsymbol{\beta}}^{(\xi)}$ for any fixed value of T:

Theorem S.1. *Let Assumptions S.1.1-S.1.3 be satisfied. Then, for $\xi \in \{L; NL\}$, as $N \to \infty$*

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}^{(\xi)} - \boldsymbol{\beta}_0\right) \Rightarrow T^{-1/2}\boldsymbol{\Sigma}_T^{(\xi)}\left(\boldsymbol{\Omega}_T^{(\xi)}\right)^{1/2}\boldsymbol{\psi} \quad (\mathcal{F}-\textit{stably}), \tag{S.8}$$

where $\boldsymbol{\Sigma}_T^{(\xi)}$, $\boldsymbol{\Omega}_T^{(\xi)}$ are $\mathcal{F}-$measurable. Both are independent of $\boldsymbol{\psi} \sim N(\mathbf{0}_D, \boldsymbol{I}_D)$.

Proof of Theorem S.1.

As in the main text, we break down the proof of this theorem into four distinct steps:

1. Establish negligibility of the bias and own terms $i = j$;
2. Derive the leading term of the asymptotic expansion;
3. Show consistency of the estimator;
4. Derive asymptotic distribution of the estimator.

To avoid notational clutter, in what follows we set $D = 1$ but we continue using vector notation unless specified otherwise. Moreover, we drop the superscript (ξ) from the estimating equations, as these are straightforwardly applicable for both linear and non-linear estimators.

Step 1. Using a similar decomposition as that in Lemma 1, we express

$$\begin{aligned}
\overline{\overline{\boldsymbol{m}}}(\boldsymbol{\beta}_0) &= \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T_1}\left(w_{j,t+1}\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t}+\lambda_i f_t\right) - w_{j,t}\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t+1}+\lambda_i f_{t+1}\right)\right) \\
&= \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{t=1}^{T_1}\left(w_{i,t+1}\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t}+\lambda_i f_t\right) - w_{i,t}\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t+1}+\lambda_i f_{t+1}\right)\right) \\
&+ \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1}\left(w_{j,t+1}\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t}+\lambda_i f_t\right) - w_{j,t}\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t+1}+\lambda_i f_{t+1}\right)\right) \\
&= \overline{\overline{\boldsymbol{m}}}^{(1)} + \overline{\overline{\boldsymbol{m}}}^{(2)} \\
&= \overline{\overline{\boldsymbol{m}}}^{(1)} - \mathrm{E}_{\mathcal{F}}[\overline{\overline{\boldsymbol{m}}}^{(1)}] + \mathrm{E}_{\mathcal{F}}[\overline{\overline{\boldsymbol{m}}}^{(1)}] + \overline{\overline{\boldsymbol{m}}}^{(2)} \\
&= \overline{\overline{\boldsymbol{m}}}^{(1)} - \mathrm{E}_{\mathcal{F}}[\overline{\overline{\boldsymbol{m}}}^{(1)}] + \overline{\overline{\boldsymbol{m}}}^{(2)} + \mathcal{O}(N^{-1}).
\end{aligned} \tag{S.9}$$

Thus, the bias term is negligible asymptotically as $N \to \infty$. Next we show that

$$\sqrt{N}(\overline{\overline{\boldsymbol{m}}}^{(1)} - \mathrm{E}_{\mathcal{F}}[\overline{\overline{\boldsymbol{m}}}^{(1)}]) = \mathrm{o}_P(1); \tag{S.10}$$

$$\sqrt{N}(\overline{\overline{\boldsymbol{m}}}^{(2)}) = \mathcal{O}_P(1). \tag{S.11}$$

For the first component observe that

$$I = \sqrt{N}(\overline{\overline{\boldsymbol{m}}}^{(1)} - \mathrm{E}_{\mathcal{F}}[\overline{\overline{\boldsymbol{m}}}^{(1)}]) = \frac{1}{\sqrt{N}}\frac{1}{N}\sum_{i=1}^{N}\overline{\boldsymbol{\mu}}_{i,T}^{(1)}. \tag{S.12}$$

Here $\mathrm{E}_{\mathcal{F}}[\overline{\boldsymbol{\mu}}_{i,T}^{(1)}] = \mathbf{0}$ by construction. By Assumption S.1.1 (b)-(c) all time-varying and time-invariant components have finite $4+\delta$ moments conditionally on $\mathcal{F}$, and also unconditionally. Repeated application of the triangle inequality and the generalized Hölder's inequality is sufficient to show that

$$\mathrm{E}_{\mathcal{F}}\left[|\overline{\boldsymbol{\mu}}_{i,T}^{(1)}|^{1+\delta}\right] < \Xi, \tag{S.13}$$

for some $\delta > 0$. Integrability of Ξ also implies that $\mathrm{E}\left[|\overline{\boldsymbol{\mu}}_{i,T}^{(1)}|^{1+\delta}\right] < \infty$, and so using Lemma 1 in Andrews (2005), we have $N^{-1}\sum_{i=1}^{N}\overline{\boldsymbol{\mu}}_{i,T}^{(1)} \xrightarrow{p} \mathbf{0}_D$. As a result,

$$\sqrt{N}(\overline{\overline{\boldsymbol{m}}}^{(1)} - \mathrm{E}_{\mathcal{F}}[\overline{\overline{\boldsymbol{m}}}^{(1)}]) = N^{-1/2}\,\mathrm{o}_P(1). \tag{S.14}$$

Step 2. Given that the first term is asymptotically negligible, we consider the second term. We expand $\overline{\overline{\boldsymbol{m}}}^{(2)}$ in the following manner:

$$\overline{\overline{\boldsymbol{m}}}^{(2)} = \sum_{s=1}^{8}\overline{\overline{\boldsymbol{m}}}^{(2.s)}, \tag{S.15}$$

where

$$\overline{\overline{\boldsymbol{m}}}^{(2.1)} = \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} c_j f_{t+1}\boldsymbol{z}_{i,t}\varepsilon_{i,t}; \qquad \overline{\overline{\boldsymbol{m}}}^{(2.2)} = \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} c_j f_{t+1}\boldsymbol{z}_{i,t}\lambda_i f_t;$$

$$\overline{\overline{\boldsymbol{m}}}^{(2.3)} = \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} \eta_{j,t+1}\boldsymbol{z}_{i,t}\varepsilon_{i,t}; \qquad \overline{\overline{\boldsymbol{m}}}^{(2.4)} = \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} \eta_{j,t+1}\boldsymbol{z}_{i,t}\lambda_i f_t;$$

$$\overline{\overline{\boldsymbol{m}}}^{(2.5)} = -\frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} c_j f_t\boldsymbol{z}_{i,t}\varepsilon_{i,t+1}; \qquad \overline{\overline{\boldsymbol{m}}}^{(2.6)} = -\frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} c_j f_t\boldsymbol{z}_{i,t}\lambda_i f_{t+1};$$

$$\overline{\overline{\boldsymbol{m}}}^{(2.7)} = -\frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} \eta_{j,t}\boldsymbol{z}_{i,t}\varepsilon_{i,t+1}; \qquad \overline{\overline{\boldsymbol{m}}}^{(2.8)} = -\frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} \eta_{j,t}\boldsymbol{z}_{i,t}\lambda_i f_{t+1}.$$

Clearly, $\overline{\overline{\boldsymbol{m}}}^{(2.2)} + \overline{\overline{\boldsymbol{m}}}^{(2.6)} = 0$. From the remaining 6 terms, we construct three terms combined:

$$\overline{\overline{\boldsymbol{m}}}^{(2.1+2.5)} = -\frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} c_j\boldsymbol{z}_{i,t}\Delta_f\varepsilon_{i,t+1}; \tag{S.16}$$

$$\overline{\overline{\boldsymbol{m}}}^{(2.3+2.7)} = \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} \boldsymbol{z}_{i,t}\left(\eta_{j,t+1}\varepsilon_{i,t} - \eta_{j,t}\varepsilon_{i,t+1}\right); \tag{S.17}$$

$$\overline{\overline{\boldsymbol{m}}}^{(2.4+2.8)} = \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} \lambda_i\boldsymbol{z}_{i,t}\Delta_f\eta_{j,t+1} = \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} \lambda_j\boldsymbol{z}_{j,t}\Delta_f\eta_{i,t+1}, \tag{S.18}$$

where $\Delta_f \eta_{j,t+1} = f_t \eta_{i,t+1} - f_{t+1}\eta_{i,t}$.

Next we consider all three terms sequentially. Firstly,

$$\begin{aligned}\overline{\overline{\boldsymbol{m}}}^{(2.1+2.5)} &= -\frac{1}{N^2 T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} c_j \boldsymbol{z}_{i,t}\Delta_f \varepsilon_{i,t+1} \\ &= -\bar{c}\frac{1}{N^2 T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} \boldsymbol{z}_{i,t}\Delta_f \varepsilon_{i,t+1} + \frac{1}{N^2 T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1} c_i \boldsymbol{z}_{i,t}\Delta_f \varepsilon_{i,t+1} \\ &= \overline{\overline{\boldsymbol{m}}}_{(1)}^{(2.1+2.5)} + \overline{\overline{\boldsymbol{m}}}_{(2)}^{(2.1+2.5)}, \end{aligned} \tag{S.19}$$

where $\bar{c} = N^{-1}\sum_{i=1}^{N} c_i$. Note that from Assumption S.1.1 each element of $c_i \boldsymbol{z}_{i,t}\Delta_f \varepsilon_{i,t+1}$ has a finite $4+\delta$ moment and it is unconditionally mean zero such that

$$\mathrm{E}_{\mathcal{F}}\left[c_i \boldsymbol{z}_{i,t}\Delta_f \varepsilon_{i,t+1}\right] = \mathrm{E}_{\mathcal{F}}\left[c_i \boldsymbol{z}_{i,t}\,\mathrm{E}_{\mathcal{F}}[\Delta_f \varepsilon_{i,t+1}|\boldsymbol{v}_i, \boldsymbol{z}_{i,t}]\right] = 0. \tag{S.20}$$

Thus, using similar arguments as before it is straightforward to show that $N\overline{\overline{\boldsymbol{m}}}_{(2)}^{(2.1+2.5)} \xrightarrow{p} \mathbf{0}_D$ from Lemma 1 in Andrews (2005). In particular,

$$\sqrt{N}\overline{\overline{\boldsymbol{m}}}_{(2)}^{(2.1+2.5)} = N^{-1/2}\,\mathrm{o}_P(1). \tag{S.21}$$

For the first component observe that

$$\begin{aligned}\sqrt{N}\overline{\overline{\boldsymbol{m}}}_{(1)}^{(2.1+2.5)} &= -\,\mathrm{E}_{\mathcal{F}}[c_i]\frac{1}{T_1\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T_1} \boldsymbol{z}_{i,t}\Delta_f \varepsilon_{i,t+1} \\ &\quad - N^{-1/2}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(c_i - \mathrm{E}_{\mathcal{F}}[c_i])\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T_1}\sum_{t=1}^{T_1} \boldsymbol{z}_{i,t}\Delta_f \varepsilon_{i,t+1}\right). \end{aligned} \tag{S.22}$$

Using Chebyshev's inequality

$$N^{-1/4}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(c_i - \mathrm{E}_{\mathcal{F}}[c_i]) = \mathrm{o}_P(1); \tag{S.23}$$

$$N^{-1/4}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T_1}\sum_{t=1}^{T_1} \boldsymbol{z}_{i,t}\Delta_f \varepsilon_{i,t+1} = \mathrm{o}_P(1). \tag{S.24}$$

Hence,

$$\sqrt{N}\overline{\overline{\boldsymbol{m}}}_{(1)}^{(2.1+2.5)} = -\,\mathrm{E}_{\mathcal{F}}[c_i]\frac{1}{T_1\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T_1} \boldsymbol{z}_{i,t}\Delta_f \varepsilon_{i,t+1} + \mathrm{o}_P(1), \tag{S.25}$$

and, consequently,

$$\sqrt{N}\overline{\overline{\boldsymbol{m}}}^{(2.1+2.5)} = -\,\mathrm{E}_{\mathcal{F}}[c_i]\frac{1}{T_1\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T_1}\boldsymbol{z}_{i,t}\Delta_f\varepsilon_{i,t+1} + \mathrm{o}_P(1). \tag{S.26}$$

The remaining two terms are more involved. Let

$$\begin{aligned}\overline{\overline{\boldsymbol{m}}}^{(2.3+2.7)} &= \frac{1}{N^2T_1}\sum_{i=2}^{N}\sum_{t=1}^{T_1}\sum_{j<i}(\boldsymbol{z}_{i,t}(\eta_{j,t+1}\varepsilon_{i,t} + \eta_{j,t}\varepsilon_{i,t+1}) - \boldsymbol{z}_{j,t}(\eta_{i,t+1}\varepsilon_{j,t} + \eta_{i,t}\varepsilon_{j,t+1})) \\ &= \frac{1}{N^2}\sum_{i=2}^{N}\boldsymbol{q}_{i,N} \\ &= \frac{1}{N^2}\sum_{i=2}^{N}\sum_{j<i}\boldsymbol{q}_{i,j}. \end{aligned} \tag{S.27}$$

It has already been established that $\mathrm{E}_{\mathcal{F}}[\overline{\overline{\boldsymbol{m}}}^{(2.3+2.7)}] = \mathbf{0}$. Moreover, $\mathrm{E}_{\mathcal{F}}[\boldsymbol{q}_{i,N}\boldsymbol{q}_{j,N}] = 0$ for all $i \neq j$. Using this result we will show that

$$\sqrt{N}\overline{\overline{\boldsymbol{m}}}^{(2.3+2.7)} = \mathrm{o}_P(1). \tag{S.28}$$

In particular, consider the variance of this term:

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{q},N} &= \mathrm{E}_{\mathcal{F}}\left[\left(\frac{1}{N}\sum_{i=2}^{N}\boldsymbol{q}_{i,N}\right)^2\right] = \frac{1}{N^2}\sum_{i=2}^{N}\mathrm{E}_{\mathcal{F}}[\boldsymbol{q}_{i,N}^2] = \frac{1}{N^2}\sum_{i=2}^{N}\mathrm{E}_{\mathcal{F}}\left[\left(\sum_{j<i}\boldsymbol{q}_{i,j}\right)^2\right] \\ &= \frac{1}{N^2}\sum_{i=2}^{N}\sum_{j<i}\mathrm{E}_{\mathcal{F}}\left[\boldsymbol{q}_{i,j}^2\right], \end{aligned} \tag{S.29}$$

where the final equality holds by conditional independence. Next, observe that the existence of finite $4+\delta$ moments implies that $\boldsymbol{\Sigma}_{\boldsymbol{q},N}$ is bounded by some Ξ. Furthermore, given that Ξ is assumed to be integrable, the above result also holds unconditionally. Direct application of Chebyshev's inequality establishes the desired result:

$$\sqrt{N}\overline{\overline{\boldsymbol{m}}}^{(2.3+2.7)} = \frac{1}{\sqrt{N}N}\sum_{i=2}^{N}\sum_{j<i}\boldsymbol{q}_{i,j} = \mathrm{o}_P(1). \tag{S.30}$$

Note that all above results extend to $D > 1$ directly as for any vector $\boldsymbol{x}$ element-wise convergence in probability implies also implies convergence of the whole vector $\boldsymbol{x}$.

Consider now $\overline{\overline{\boldsymbol{m}}}^{(2.4+2.8)}$. We have

$$\begin{aligned}\boldsymbol{m}^{(2.4+2.8)} &= \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1}\lambda_j \boldsymbol{z}_{j,t}\Delta_f\eta_{i,t+1} \\ &= \frac{N_1}{N^2T_1}\sum_{i=1}^{N}\sum_{t=1}^{T_1}\mathrm{E}_{\mathcal{F}}[\boldsymbol{z}_{i,t}\lambda_j]\Delta_f\eta_{i,t+1} + \frac{1}{N^2T_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{t=1}^{T_1}\Delta\boldsymbol{g}_{(\boldsymbol{z}),j,t}\Delta_f\eta_{i,t+1},\end{aligned} \tag{S.31}$$

where

$$\Delta\boldsymbol{g}_{(\boldsymbol{z}),j,t} = \boldsymbol{z}_{j,t}\lambda_j - \mathrm{E}_{\mathcal{F}}[\boldsymbol{z}_{j,t}\lambda_j].$$

Since the second component has mean-zero, using identical arguments to those employed for $\overline{\overline{\boldsymbol{m}}}_{N,T}^{(2.3+2.7)}$, it follows that

$$\begin{aligned}\sqrt{N}\overline{\overline{\boldsymbol{m}}}^{(2.4+2.8)} &= \frac{N_1}{N}\frac{1}{T_1\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T_1}\mathrm{E}_{\mathcal{F}}[\boldsymbol{z}_{i,t}\lambda_j]\Delta_f\eta_{i,t+1} + \mathrm{o}_P(1) \\ &= \frac{1}{T_1\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T_1}\mathrm{E}_{\mathcal{F}}[\boldsymbol{z}_{i,t}\lambda_j]\Delta_f\eta_{i,t+1} + \mathrm{o}_P(1).\end{aligned} \tag{S.32}$$

Collecting all terms:

$$\begin{aligned}\sqrt{N}\overline{\overline{\boldsymbol{m}}}(\boldsymbol{\beta}_0) =& \,\mathrm{E}_{\mathcal{F}}[c_i]\frac{1}{T_1\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T_1}\boldsymbol{z}_{i,t}\left(f_{t+1}\varepsilon_{i,t} - f_t\varepsilon_{i,t+1}\right) \\ &- \frac{1}{T_1\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T_1}\mathrm{E}_{\mathcal{F}}[\boldsymbol{z}_{i,t}\lambda_i]\left(f_{t+1}\eta_{i,t} - f_t\eta_{i,t+1}\right) \\ &+ \mathrm{o}_P(1).\end{aligned} \tag{S.33}$$

At first we establish $\mathcal{F}-$stable convergence of the leading term in $\sqrt{N}\overline{\overline{\boldsymbol{m}}}_{N,T}(\boldsymbol{\beta}_0)$. In particular, let

$$\begin{aligned}\sqrt{N}\overline{\overline{\boldsymbol{m}}}(\boldsymbol{\beta}_0) =& \,\mathrm{E}_{\mathcal{F}}[c_i]\frac{1}{T_1\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T_1}\boldsymbol{z}_{i,t}\left(f_{t+1}\varepsilon_{i,t} - f_t\varepsilon_{i,t+1}\right) \\ &- \frac{1}{T_1\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T_1}\mathrm{E}_{\mathcal{F}}[\boldsymbol{z}_{i,t}\lambda_i]\left(f_{t+1}\eta_{i,t} - f_t\eta_{i,t+1}\right) + \mathrm{o}_P(1) \\ =& \,\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{\boldsymbol{\mu}}_{i,T}^{(\xi)} + \mathrm{o}_P(1),\end{aligned} \tag{S.34}$$

where $\overline{\boldsymbol{\mu}}_{i,T}^{(\xi)}$ is defined implicitly. Next denote by $\mathcal{C}_i = \sigma\left(\mathcal{F} \vee \{\boldsymbol{\Upsilon}_j\}_{j=1}^{i}\right)$ the σ-field generated by $\mathcal{F}$ and $(\boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_i)$. Then $\{\overline{\boldsymbol{\mu}}_{i,T}^{(\xi)}, \mathcal{C}_i : i \geq 1\}$ is a Martingale Difference

sequence (element-wise) since by Assumption S.1.1 (a), the data are i.i.d. conditionally on $\mathcal{F}$. Also let

$$\boldsymbol{\Omega}_T^{(\xi)} = T_1 N^{-1} \sum_{i=1}^{N} \mathrm{E}\left[\overline{\boldsymbol{\mu}}_{i,T}^{(\xi)} \left(\overline{\boldsymbol{\mu}}_{i,T}^{(\xi)}\right)' | \mathcal{C}_{i-1}\right] = T_1 N^{-1} \sum_{i=1}^{N} \mathrm{E}\left[\overline{\boldsymbol{\mu}}_{i,T}^{(\xi)} \left(\overline{\boldsymbol{\mu}}_{i,T}^{(\xi)}\right)' | \mathcal{F}\right] = T_1 \, \mathrm{E}\left[\overline{\boldsymbol{\mu}}_{i,T}^{(\xi)} \left(\overline{\boldsymbol{\mu}}_{i,T}^{(\xi)}\right)' | \mathcal{F}\right]. \tag{S.35}$$

The last equality follows by conditional i.i.d. in Assumption S.1.1 (a). Finally, using Theorem 3.2. and Corollary 3.1 in Hall and Heyde (1980) in conjunction with the Cramér-Wold device we obtain

$$\sqrt{N}\overline{\overline{\boldsymbol{m}}}(\boldsymbol{\beta}_0) \Rightarrow T_1^{-1/2} \left(\boldsymbol{\Omega}_T^{(\xi)}\right)^{1/2} \boldsymbol{\psi} \quad \text{(stably)}, \tag{S.36}$$

where $\boldsymbol{\psi} \sim N(\mathbf{0}_D, \boldsymbol{I}_D)$. This result holds provided that each element $\zeta_{i,t}$ of $\overline{\boldsymbol{\mu}}_{i,T}^{(\xi)}$ satisfies the conditional Lindeberg's condition:

$$N^{-1} \sum_{i=1}^{N} \mathrm{E}_{\mathcal{F}}[\zeta_{i,T}^2 I(|\zeta_{i,T}| > \sqrt{N}\varepsilon)], \quad \text{for all } \varepsilon > 0. \tag{S.37}$$

Given that the conditional Lyapunov's condition implies the conditional Lindeberg's condition, it is sufficient that $\mathrm{E}_{\mathcal{F}}[|\zeta_{i,T}|^{2+\delta}] < \infty$ for some $\delta > 0$ which in our case can be verified using Hölder's inequality.

Step 3. The proof of consistency is fairly standard along the lines of Newey and McFadden (1994). Sufficient conditions are satisfied given the existence of finite moments due to Assumption S.1.1 and the global identification over a compact Θ by Assumption 3.3.

Step 4. Asymptotic distribution can be obtained by expanding the first-order conditions around the true value (again, we omit the ξ index):

$$\boldsymbol{W}_{N,T}\overline{\overline{\boldsymbol{m}}}(\hat{\boldsymbol{\beta}}) = \boldsymbol{W}_{N,T} \left[\overline{\overline{\boldsymbol{m}}}(\boldsymbol{\beta}_0) + \left(\overline{\overline{\boldsymbol{\Gamma}}}_T + \frac{1}{2}\sum_{k=1}^{K} \overline{\overline{\boldsymbol{H}}}_{T,k}(\hat{\beta}_k - \beta_{0,k})\right)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right]. \tag{S.38}$$

Our assumptions ensure that the following limits

$$\operatorname*{plim}_{N\to\infty} \overline{\overline{\boldsymbol{\Gamma}}}_T(\boldsymbol{\beta}_0); \quad \operatorname*{plim}_{N\to\infty} \overline{\overline{\boldsymbol{H}}}_{T,k}(\boldsymbol{\beta}_0)$$

are finite and $\mathcal{F}$ measurable. Given that convergence in probability implies convergence in distribution, the remainder of the proof follows directly from Proposition A.2 in Kuersteiner and Prucha (2013) and an application of the continuous mapping theorem.

□

Finally, notice that since matrices $\boldsymbol{\Sigma}_T^{(\xi)}$ and $\boldsymbol{\Omega}_T^{(\xi)}$ are random (but independent of $\boldsymbol{\psi}$), the unconditional limiting distribution of the proposed estimator is mixed-normal rather than normal. While this is an important distinction when discussing the properties of the estimator itself compared to existing literature (e.g. Ahn et al. 2013 and Robertson and Sarafidis 2015), it plays no role for inference procedures that make use of standardized (pivotal) statistics. This is so long as both $\boldsymbol{\Omega}_T^{(\xi)}$ and $\boldsymbol{\Sigma}_T^{(\xi)}$ can be consistently estimated from their sample analogues.

S.2. Elimination of Common Time Effects

Consider the following $[D_t \times 1]$ vector of estimating equations available at time period t, expressed in terms of deviations from time-specific cross-sectional averages:

$$\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)}(\boldsymbol{\beta}_0) = \frac{1}{NN_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[w_{j,t,t+1}^{(\xi)}\underline{\boldsymbol{z}}_{i,t}\left(\underline{\varepsilon}_{i,t}+\underline{\lambda}_i f_t\right) - w_{j,t,t}^{(\xi)}\underline{\boldsymbol{z}}_{i,t}\left(\underline{\varepsilon}_{i,t+1}+\underline{\lambda}_i f_{t+1}\right)\right]. \tag{S.39}$$

for $t = 1, \ldots, T-1$, where $\underline{\boldsymbol{z}}_{i,t} = \boldsymbol{z}_{i,t} - \overline{\boldsymbol{z}}_t$ with $\overline{\boldsymbol{z}}_t = N^{-1}\sum_{i=1}^{N}\boldsymbol{z}_{i,t}$, and similarly for the remaining variables. The following lemma shows that the averaged estimating equations, transformed in terms of deviations from time-specific cross-sectional averages, can be biased for T large.

Lemma S.1. *Suppose that Assumption 3.1 is satisfied. Then for all $t = 1, \ldots, T-1$,*

$$\mathrm{E}_{\mathcal{D}}\left[\widetilde{\boldsymbol{m}}_t^{(L)(TE)}(\boldsymbol{\beta}_0)\right] = \mathcal{O}_P(N^{-2}); \tag{S.40}$$

$$\mathrm{E}_{\mathcal{D}}\left[\widetilde{\boldsymbol{m}}_t^{(NL)(TE)}(\boldsymbol{\beta}_0)\right] = \mathbf{0}_{D_t}. \tag{S.41}$$

Proof. See below. □

As a result of the lemma above, we have

$$\sqrt{NT}\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}_{\mathcal{D}}\left[\widetilde{\boldsymbol{m}}_t^{(L)(TE)}(\boldsymbol{\beta}_0)\right] = \mathcal{O}_P\left(\sqrt{\frac{T}{N^3}}\right) = \mathrm{o}_P(1). \tag{S.42}$$

Thus, provided that $T/N^3 \to 0$, which is likely to be satisfied in most relevant applications, this bias term for the transformed linear estimating equations is negligible.

Proof of Lemma S.1.

Let $\Delta \boldsymbol{z}_{i,t} = \boldsymbol{z}_{i,t} - \mathrm{E}_{\mathcal{F}}[\boldsymbol{z}_{i,t}]$ and similarly for the remaining variables. Noting that $\Delta\varepsilon_{i,t} = \varepsilon_{i,t}$ because $\mathrm{E}_{\mathcal{F}}[\varepsilon_{i,t}] = 0$, we have

$$\begin{aligned}
\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)}(\boldsymbol{\beta}_0) &= \frac{1}{NN_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[w_{j,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t}+\Delta\lambda_i f_t\right) - w_{j,t,t}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t+1}+\Delta\lambda_i f_{t+1}\right)\right] \\
&\quad - \frac{1}{NN_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[w_{j,t,t+1}^{(\xi)}\overline{\Delta\boldsymbol{z}}_t\left(\varepsilon_{i,t}+\Delta\lambda_i f_t\right) - w_{j,t,t}^{(\xi)}\overline{\Delta\boldsymbol{z}}_t\left(\varepsilon_{i,t+1}+\Delta\lambda_i f_{t+1}\right)\right] \\
&\quad - \frac{1}{NN_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[w_{j,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\overline{\varepsilon}_t+\overline{\Delta\lambda}f_t\right) - w_{j,t,t}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\overline{\varepsilon}_t+\overline{\Delta\lambda}f_{t+1}\right)\right] \\
&\quad + \frac{1}{NN_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[w_{j,t,t+1}^{(\xi)}\overline{\Delta\boldsymbol{z}}_t\left(\overline{\varepsilon}_t+\overline{\Delta\lambda}f_t\right) - w_{j,t,t}^{(\xi)}\overline{\Delta\boldsymbol{z}}_t\left(\overline{\varepsilon}_t+\overline{\Delta\lambda}f_{t+1}\right)\right] \\
&= \widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(1)} - \widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(2)} - \widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(3)} + \widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(4)}. \qquad \text{(S.43)}
\end{aligned}$$

As in Lemma 1 it is easy to show that $\mathrm{E}_{\mathcal{D}}[\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(1)}] = \mathbf{0}_{D_t}$. Consider all remaining components individually:

$$\begin{aligned}
\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(2)} &= \frac{1}{N^2N_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[w_{j,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t}+\Delta\lambda_i f_t\right) - w_{j,t,t}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t+1}+\Delta\lambda_i f_{t+1}\right)\right] \\
&\quad + \frac{1}{N^2N_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[w_{j,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{j,t}\left(\varepsilon_{i,t}+\Delta\lambda_i f_t\right) - w_{j,t,t}^{(\xi)}\Delta\boldsymbol{z}_{j,t}\left(\varepsilon_{i,t+1}+\Delta\lambda_i f_{t+1}\right)\right] \\
&\quad + \frac{1}{N^2N_1}\sum_{i=1}^{N}\sum_{k\neq i}^{N}\sum_{j\neq i,j\neq k}^{N}\left[w_{j,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{k,t}\left(\varepsilon_{i,t}+\Delta\lambda_i f_t\right) - w_{j,t,t}^{(\xi)}\Delta\boldsymbol{z}_{k,t}\left(\varepsilon_{i,t+1}+\Delta\lambda_i f_{t+1}\right)\right]. \qquad \text{(S.44)}
\end{aligned}$$

Using Lemma 1 it is easy to show that only the second component has non-zero conditional expectations, in particular

$$\mathrm{E}_{\mathcal{D}}\left[\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(2)}\right] = \frac{1}{N^2N_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[\mathrm{E}_{\mathcal{D}}[q_{j,t}\varepsilon_{j,t+1}^{(\xi)}\Delta\boldsymbol{z}_{j,t}]\left(\Delta\lambda_i f_t\right) - \mathrm{E}_{\mathcal{D}}[q_{j,t}\varepsilon_{j,t}^{(\xi)}\Delta\boldsymbol{z}_{j,t}]\left(\Delta\lambda_i f_{t+1}\right)\right]. \qquad \text{(S.45)}$$

However, for $\xi = NL$ the above conditional expectations are zero. Next,

$$\begin{aligned}\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(3)} &= \frac{1}{N^2N_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[w_{j,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t}+\Delta\lambda_i f_t\right) - w_{j,t,t}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t+1}+\Delta\lambda_i f_{t+1}\right)\right] \\ &+ \frac{1}{N^2N_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[w_{j,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{j,t}+\Delta\lambda_j f_t\right) - w_{j,t,t}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{j,t+1}+\Delta\lambda_j f_{t+1}\right)\right] \\ &+ \frac{1}{N^2N_1}\sum_{i=1}^{N}\sum_{k\neq i}^{N}\sum_{j\neq i,j\neq k}^{N}\left[w_{j,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{k,t}+\Delta\lambda_k f_t\right) - w_{j,t,t}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{k,t+1}+\Delta\lambda_k f_{t+1}\right)\right].\end{aligned} \tag{S.46}$$

Here

$$\mathrm{E}_{\mathcal{D}}\left[\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(3)}\right] = \frac{1}{N^2N_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[\mathrm{E}_{\mathcal{D}}[q_{j,t}\varepsilon_{j,t+1}^{(\xi)}\varepsilon_{j,t}] - \mathrm{E}_{\mathcal{D}}[q_{j,t}\varepsilon_{j,t}^{(\xi)}\varepsilon_{j,t+1}]\left(\Delta\lambda_j f_{t+1}\right)\right]\mathrm{E}_{\mathcal{D}}[\Delta\boldsymbol{z}_{i,t}]. \tag{S.47}$$

For the final component note that

$$\begin{aligned}\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(4)} &= \frac{1}{N^2N_1}\sum_{i=1}^{N}\left[w_{i,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t}+\Delta\lambda_i f_t\right) - w_{i,t,t}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t+1}+\Delta\lambda_i f_{t+1}\right)\right] \\ &+ \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[w_{j,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{j,t}+\Delta\lambda_j f_t\right) - w_{j,t,t}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{j,t+1}+\Delta\lambda_j f_{t+1}\right)\right] \\ &+ \frac{N_1}{N}\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)(2)}.\end{aligned} \tag{S.48}$$

Based on the above decomposition it is not difficult to show that

$$\begin{aligned}\mathrm{E}_{\mathcal{D}}\left[\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)}(\boldsymbol{\beta}_0)\right] &= \frac{1}{N^2}\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}_{\mathcal{D}}\left[w_{i,t,t+1}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t}+\Delta\lambda_i f_t\right) - w_{i,t,t}^{(\xi)}\Delta\boldsymbol{z}_{i,t}\left(\varepsilon_{i,t+1}+\Delta\lambda_i f_{t+1}\right)\right] \\ &- \frac{1}{N^2}\frac{1}{NN_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[\mathrm{E}_{\mathcal{D}}[q_{j,t}\varepsilon_{j,t+1}^{(\xi)}\Delta\boldsymbol{z}_{j,t}]\left(\Delta\lambda_i f_t\right) - \mathrm{E}_{\mathcal{D}}[q_{j,t}\varepsilon_{j,t}^{(\xi)}\Delta\boldsymbol{z}_{j,t}]\left(\Delta\lambda_i f_{t+1}\right)\right] \\ &- \frac{1}{N^2}\frac{1}{NN_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[\mathrm{E}_{\mathcal{D}}[q_{j,t}\varepsilon_{j,t+1}^{(\xi)}\varepsilon_{j,t}] - \mathrm{E}_{\mathcal{D}}[q_{j,t}\varepsilon_{j,t}^{(\xi)}\varepsilon_{j,t+1}]\right]\mathrm{E}_{\mathcal{D}}[\Delta\boldsymbol{z}_{i,t}].\end{aligned} \tag{S.49}$$

Using steps similar to those in Lemma 1, it is straightforward to show that

$$\mathrm{E}_{\mathcal{D}}\left[\widetilde{\boldsymbol{m}}_t^{(\xi)(TE)}(\boldsymbol{\beta}_0)\right] = \mathcal{O}_P(N^{-2}). \tag{S.50}$$

□

Notice that for $\xi = NL$ the expression above is exactly zero:

$$\mathrm{E}_{\mathcal{D}}\left[\widetilde{\boldsymbol{m}}_t^{(NL)(TE)}(\boldsymbol{\beta}_0)\right] = -\frac{1}{N^2}\frac{1}{NN_1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left[\mathrm{E}_{\mathcal{D}}[q_{j,t}\varepsilon_{j,t+1}\varepsilon_{j,t}] - \mathrm{E}_{\mathcal{D}}[q_{j,t}\varepsilon_{j,t}\varepsilon_{j,t+1}]\right]\mathrm{E}_{\mathcal{D}}[\Delta \boldsymbol{z}_{i,t}] = \mathbf{0}_{D_t}.$$

S.3. Identification-Robust Inference

In Section 5.2 of the main text we discussed potential failures of local and global identification for the methods proposed in this paper. If identification fails, then none of the standard testing procedures can be used. Instead, one needs to use a more robust approach to inference.

An intermediate approach is to use two-step identification robust confidence sets proposed by Andrews (2018). Depending on the identification strength, this approach constructs confidence intervals that closely resemble those based on non-robust or robust tests. Alternatively, one can employ (full) identification robust inference procedures, thus avoiding an assumption in regards to potential identification strength. In what follows we will briefly discuss readily available procedures and their pitfalls in our setup. Building upon the idea of Anderson and Rubin (1949), Stock and Wright (2000) introduced the nonlinear Anderson-Rubin (AR) test, commonly known as the "S" test. By construction, this test is robust to identification failure and weak identification. The S statistic in our setup is given by:

$$AR_{N,T}(\boldsymbol{\beta}) = N(T_1)\left(\overline{\widetilde{\boldsymbol{m}}}^{(\xi)}(\boldsymbol{\beta})\right)'\left[\widehat{\boldsymbol{\Omega}}^{(\xi)}(\boldsymbol{\beta})\right]^{-1}\overline{\widetilde{\boldsymbol{m}}}^{(\xi)}(\boldsymbol{\beta}), \tag{S.51}$$

where $\widehat{\boldsymbol{\Omega}}^{(\xi)}(\boldsymbol{\beta})$ is defined in Section 5.1 of the main text. Irrespective of identification strength and potential failure of global identification, under $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ we have

$$AR_{N,T}(\boldsymbol{\beta}_0) \xrightarrow{d} \chi^2_D. \tag{S.52}$$

Thus, confidence intervals can be constructed by inverting the S statistic. The performance of this method in finite samples is illustrated in Section S.4.2.

Remark 1. The S test may have deficient power in overidentified settings where $D >> K$, in which case one can consider the so-called K test of Kleibergen (2005). However in the non-linear case, due to the multi-modal nature of the objective function, the K test statistic may also have low power over a sizeable part of the parameter space, as all LM statistics generally lack power around local minima and turning points. Alternatively, since neither S or K statistics are efficient, one could consider the GMM extensions of the Moreira (2003) CLR statistic, as put forward by Andrews and Mikusheva (2016) and Andrews (2016).

To this date we are not aware of any theoretical or simulation evidence on the use of identification-robust testing procedures for GMM methods employed in fixed T factor-augmented panels.[1] This may be partially due to the longstanding complication of identification-robust GMM inference when it comes to subset inference. This problem will be especially pronounced for all fixed T GMM procedures currently available, as one needs to estimate a large set of nuisance parameters, that is, either $\boldsymbol{F}$ in the case of Ahn et al. (2013) and/or $\boldsymbol{G}$ for Robertson and Sarafidis (2015), while at the same time use many instruments. As a result, confidence intervals constructed using (say) the projection method of Dufour and Taamouti (2005) and Chaudhuri and Zivot (2011) are not only difficult to obtain, but they may suffer from low power. The fact that our procedure does not require estimation of any nuisance parameters, substantially simplifies identification-robust inference.

S.4. Identification in the Panel AR(1) Model using Averaged Moment Conditions

S.4.1. Theoretical analysis

We consider a reduced form of the panel AR(1) model with a single factor:

$$y_{i,t} = \lambda_i g_t + u_{i,t}; \quad u_{i,t} = \alpha u_{i,t-1} + \varepsilon_{i,t}; \quad g_t = \alpha g_{t-1} + f_t; \quad t = 1, \ldots, T, \tag{S.53}$$

where $|\alpha| < 1$. Without loss of generality, we focus on the nonlinear estimator. Setting $z_{i,t} = y_{i,t-1}$ and $q_{i,t} = y_{i,t-2}$, the expected Jacobian matrix at time t takes the following form:

$$\begin{aligned} \boldsymbol{\Gamma}_t^{(NL)} = {} & \mathrm{E}_{\mathcal{F}}[\lambda_i^2] \left(g_{t-2} f_t \, \mathrm{E}_{\mathcal{F}}[u_{i,t-1} u_{i,t}] + g_{t-1} f_{t+1} \, \mathrm{E}_{\mathcal{F}}[u_{i,t-1} u_{i,t-2}] \right) \\ & - \mathrm{E}_{\mathcal{F}}[\lambda_i^2] \left(g_{t-2} f_{t+1} \, \mathrm{E}_{\mathcal{F}} \left[u_{i,t-1}^2 \right] + g_{t-1} f_t \, \mathrm{E}_{\mathcal{F}}[u_{i,t} u_{i,t-2}] \right). \end{aligned} \tag{S.54}$$

Thus, for T fixed, local identification is guaranteed so long as the factors are stochastic with a continuous distribution (or non-zero constants). For T large, under suitable stationarity conditions the Jacobian matrix corresponding to the averaged estimating equations, converges in probability to

$$\boldsymbol{\Gamma}^{(NL)} = \mathrm{E}_{\mathcal{F}}[\lambda_i^2] \frac{\sigma_\varepsilon^2}{1 - \alpha^2} \left(\alpha \, \mathrm{E}[g_{t-2} f_t] - \mathrm{E}[g_{t-2} f_{t+1}] - \alpha^2 \, \mathrm{E}[g_{t-1} f_t] \right). \tag{S.55}$$

[1] Properties of identification robust procedures are only discussed in the context of the simple panel AR(1) model with fixed effects; see Bun and Kleibergen (2016) and Dovonon et al. (2020).

Thus in this case, assuming $\mathrm{E}[f_t] = 0$, local identification requires that the factor is serially correlated. On the other hand, for the stacked estimating equations it is straightforward to see that identification relies purely on fourth-order cross-moments of unobserved common factors and serial correlation is not required. Similar results apply to the linear approach and therefore we refrain from providing any details.

In summary, for the estimator that makes use of averaged estimating equations, provided certain "fixed T" identification restrictions are satisfied, both linear and non-linear GMM estimators have desirable asymptotic properties as $N \to \infty$, even in cases where the expected Jacobian matrix of the moment conditions becomes singular as $T \to \infty$. As a byproduct, we note that a simple comparison between Eq. (S.54) and (S.55) reveals that within the proposed framework, identification of autoregressive models with a unit root is feasible under "fixed T" asymptotics even if it fails when $T \to \infty$. This principle also underlies the approach put forward by Robertson et al. (2018). Essentially, drawing a parallel with existing results on fixed effects models, when T is fixed the presence of unobserved factors renders the process of $y_{i,t}$ non mean-stationary by construction, which may help to increase the signal of the model.[2]

We conclude this section by studying global identification for the AR model using averaged estimating equations. It is demonstrated that the use of *time-invariant* weights alone can be problematic, even if multiple weights are employed.

Proposition S.1. *Consider the panel AR(1) model in Eq.* (S.53) *and let* $\boldsymbol{z}_{i,t} = (y_{i,t-1}, y_{i,t-2})'$. *Suppose that* $\mathrm{E}_{\mathcal{F}}[q_i \lambda_i] \neq 0$. *Then*

(a) *Given* $\boldsymbol{q}_i = (q_i^{(1)}, q_i^{(2)})'$ *(multiple time-invariant weights), Assumption 3.3 is satisfied for T fixed, so long as f_t has a continuous distribution.*

(b) *As $T \to \infty$, global identification fails for the same setup as in (a). The pseudo true value is given by $\alpha_* = 1$.*

(c) *Given* $\boldsymbol{q}_{i,t} = (q_i^{(1)}, y_{i,t-2})'$, *Assumption 3.3 is satisfied for both T fixed and $T \to \infty$ cases, so long as f_t has a continuous distribution (for T fixed), or is serially correlated (for T large).*

Proof of Proposition S.1.

At first we consider the case where only a time-invariant weight is used. Furthermore, for simplicity we assume that the process $u_{i,t}$ is covariance-stationary and independent

[2]See e.g. Hayakawa (2009) and Bun and Sarafidis (2015).

of factors, such that[3]

$$\gamma(s) = \mathrm{E}[u_{i,t}u_{i,t-s}] = \sigma_\varepsilon^2 \alpha_0^{|s|}(1-\alpha_0^2)^{-1}. \tag{S.56}$$

In this case, the expected Jacobian matrix reduces to

$$\begin{aligned}
\boldsymbol{\Gamma}_T^{(NL)} &= \frac{1}{T_1}\,\mathrm{E}_{\mathcal{F}}[q_i\lambda_i] \begin{pmatrix} \sum_{t=1}^{T_1}\left(f_t\,\mathrm{E}[u_{i,t-1}u_{i,t}] - f_{t+1}\,\mathrm{E}[u_{i,t-1}^2]\right) \\ \sum_{t=2}^{T_1}\left(f_t\,\mathrm{E}[u_{i,t-2}u_{i,t}] - f_{t+1}\,\mathrm{E}[u_{i,t-1}u_{i,t-2}]\right)] \end{pmatrix} \\
&= \frac{\mathrm{E}[u_{i,t}^2]}{T_1}\,\mathrm{E}_{\mathcal{F}}[q_i\lambda_i] \begin{pmatrix} \sum_{t=1}^{T_1}\left(f_t\alpha_0 - f_{t+1}\right) \\ \alpha_0\sum_{t=2}^{T_1}\left(f_t\alpha_0 - f_{t+1}\right). \end{pmatrix}
\end{aligned}$$

On the other hand, the matrix of expected second derivatives is given by

$$\begin{aligned}
\boldsymbol{H}_T^{(NL)} &= \frac{2}{T_1}\,\mathrm{E}_{\mathcal{F}}[q_i\lambda_i] \begin{pmatrix} \sum_{t=1}^{T_1}\left(g_t\,\mathrm{E}[u_{i,t-1}^2] - g_{t-1}\,\mathrm{E}[u_{i,t-1}u_t]\right) \\ \sum_{t=2}^{T_1}\left(g_t\,\mathrm{E}[u_{i,t-1}u_{i,t-2}] - g_{t-1}\,\mathrm{E}[u_{i,t-2}u_t]\right) \end{pmatrix} \\
&= \frac{2\,\mathrm{E}[u_{i,t}^2]}{T_1}\,\mathrm{E}_{\mathcal{F}}[q_i\lambda_i] \begin{pmatrix} \sum_{t=1}^{T_1}\left(g_t - g_{t-1}\alpha_0\right) \\ \alpha_0\sum_{t=2}^{T_1}\left(g_t - g_{t-1}\alpha_0\right) \end{pmatrix} \\
&= \frac{2\,\mathrm{E}[u_{i,t}^2]}{T_1}\,\mathrm{E}_{\mathcal{F}}[q_i\lambda_i] \begin{pmatrix} \sum_{t=1}^{T_1} f_t \\ \alpha_0\sum_{t=2}^{T_1} f_t \end{pmatrix}.
\end{aligned}$$

Note that because of different summation (i.e. $\sum_{t=1}^{T_1}$ vs $\sum_{t=2}^{T_1}$), the pseudo true-values for each instrument are different. In particular, for $z_{i,t} = y_{i,t-1}$ the pseudo true value is:

$$\alpha_T^* = \alpha_0 - \frac{\sum_{t=1}^{T_1}(\alpha_0 f_t - f_{t+1})}{\sum_{t=1}^{T_1} f_t} = \frac{\sum_{t=1}^{T_1} f_{t+1}}{\sum_{t=1}^{T_1} f_t} = 1 + \mathcal{O}_P(T^{-1/2}). \tag{S.57}$$

while for $z_{i,t} = y_{i,t-2}$ the corresponding pseudo-true value is

$$\alpha_T^{**} = \alpha_0 - \frac{\sum_{t=2}^{T_1}(\alpha_0 f_t - f_{t+1})}{\sum_{t=2}^{T_1} f_t} = \frac{\sum_{t=2}^{T_1} f_{t+1}}{\sum_{t=2}^{T_1} f_t} \neq \alpha_T^*. \tag{S.58}$$

However, since the difference between α_T^* and α_T^{**} is of order $\mathcal{O}_P(T^{-1})$ for T large the model is not globally identified, with the pseudo-true value of $\mathrm{plim}_{T\to\infty}\,\alpha_T^* = \mathrm{plim}_{T\to\infty}\,\alpha_T^{**} = 1$.

Finally, we consider the setup with one additional time-varying weight, $q_{i,t} = y_{i,t-2}$. In this setup, we only use $z_{i,t} = y_{i,t-1}$ as an instrument. For the combination with

[3]All results will stay the same if we assume a finite past initialization. Actually, in that case, the global identification condition is more likely to be satisfied for T fixed.

time-invariant weight we showed previously that the pseudo-true value satisfies

$$\alpha_T^* = \alpha_0 - \frac{\sum_{t=1}^{T_1}(\alpha_0 f_t - f_{t+1})}{\sum_{t=1}^{T_1} f_t} = \frac{\sum_{t=1}^{T_1} f_{t+1}}{\sum_{t=1}^{T_1} f_t} = 1 + \mathcal{O}_P(T^{-1/2}). \tag{S.59}$$

The Jacobian matrix for the time-varying weight is given by

$$\begin{aligned}
\boldsymbol{\Gamma}_T^{(NL)} &= \mathrm{E}_{\mathcal{F}}[\lambda_i^2]\frac{1}{T_1}\sum_{t=1}^{T_1} g_{t-2}\left(f_t(\gamma(1) + \mathrm{E}_{\mathcal{F}}[\lambda_i^2]g_{t-1}g_t) - f_{t+1}(\gamma(0) + \mathrm{E}_{\mathcal{F}}[\lambda_i^2]g_{t-1}^2)\right) \\
&\quad + \mathrm{E}_{\mathcal{F}}[\lambda_i^2]\frac{1}{T_1}\sum_{t=1}^{T_1} g_{t-1}\left(f_{t+1}(\gamma(1) + \mathrm{E}_{\mathcal{F}}[\lambda_i^2]g_{t-1}g_{t-2}) - f_t(\gamma(2) + \mathrm{E}_{\mathcal{F}}[\lambda_i^2]g_{t-2}g_t)\right) \\
&= -\mathrm{E}_{\mathcal{F}}[\lambda_i^2]\gamma(0)\frac{1}{T_1}\sum_{t=1}^{T_1}(f_{t+1} - \alpha_0 f_t)(g_{t-2} - \alpha_0 g_{t-1}).
\end{aligned} \tag{S.60}$$

After similar derivations one can show that the expected second derivative is of the form

$$\boldsymbol{H}_T^{(NL)} = 2\,\mathrm{E}_{\mathcal{F}}[\lambda_i^2]\gamma(0)\frac{1}{T_1}\sum_{t=1}^{T_1} f_t(g_{t-2} - \alpha_0 g_{t-1}). \tag{S.61}$$

Thus the pseudo-true value for this setup is given by:

$$\alpha^\dagger = \frac{\sum_{t=1}^{T_1} f_{t+1}(g_{t-2} - \alpha_0 g_{t-1})}{\sum_{t=1}^{T_1} f_t(g_{t-2} - \alpha_0 g_{t-1})} = \frac{\mu_f^2 + \sum_{t=1}^{T_1}\widetilde{f}_{t+1}(\widetilde{g}_{t-2} - \alpha_0\widetilde{g}_{t-1})}{\mu_f^2 + \sum_{t=1}^{T_1}\widetilde{f}_t(\widetilde{g}_{t-2} - \alpha_0\widetilde{g}_{t-1})}, \tag{S.62}$$

where $\widetilde{f}_t$ and $\widetilde{g}_t$ denote the demeaned version of f_t and g_t, respectively, and $\mu_f = \mathrm{E}[f_t]$. From the above expression we can see that $\alpha^\dagger$ clearly differs from α^*, both for T fixed and $T \to \infty$.

For example, if f_t is an MA(1) with parameter θ and increments ν_t, then

$$\operatorname*{plim}_{T\to\infty} \alpha^\dagger = \frac{\mu_f^2}{\mu_f^2 - \alpha_0\theta\sigma_\nu^2}, \tag{S.63}$$

which equals 1 only if either $\theta = 0$ (no autocorrelation in f_t) or $\alpha_0 = 0$. On the other hand, if f_t is an AR(1) with parameter α_f:

$$\operatorname*{plim}_{T\to\infty} \alpha^\dagger = \frac{\mu_f^2 + \alpha_f^2(1 - \alpha_0\alpha_f^{-1}(\alpha_f - 1)}{\mu_f^2 + \alpha_f(1 - \alpha_0\alpha_f)^{-1}(\alpha_f - 1)}, \tag{S.64}$$

which, again, equals 1 only if $\alpha_f = 0$ or $\mu_f = 0$ and $\alpha_f = 1$. □

The above proposition shows that in the panel AR(1) model, global identification cannot be guaranteed by using multiple time-invariant weights alone. That is, the use

of at least one time-varying weight is necessary for large T identification. Overall, the previous discussion indicates that the panel AR(1) model proxies the worst case scenario in terms of local and global identification. In any case, we do advise to use time-varying weights for more general models, especially with predetermined regressors.

S.4.2. Simulations

In order to illustrate how theoretical results on identification of the panel AR model are relevant in finite samples, we consider the following DGP:

$$y_{i,t} = \alpha y_{i,t-1} + \lambda_i f_t + \varepsilon_{i,t}; \quad d_{i,t} = \lambda_i^d f_t + \varepsilon_{i,t}^d; \tag{S.65}$$

$$f_t = \mu_f + f_t^*; \quad f_t^* = \alpha_f f_{t-1}^* + \sqrt{1-\alpha_f^2} u_t, \tag{S.66}$$

for $t = -4, \dots T$. This is a simplified version of the model considered in Juodis and Sarafidis (2018), as well as that used in the main text. All time-varying error terms are mutually independent $N(0,1)$. All dynamic processes are initialized in the recent past such that

$$y_{i,-5} = \lambda_i f_{-5} + \varepsilon_{i,-5}; \quad d_{i,-5} = \lambda_i^d f_{-5} + \varepsilon_{i,-5}^{(d)}; \quad f_{-5} = \mu_f + u_{-5}. \tag{S.67}$$

We specify $\mathrm{E}[\lambda_i] = \mathrm{E}[\lambda_i^d] = 1$, $\alpha_f = 0.5$, $\alpha = \{0.4; 0.8\}$, whereas $\mu_f = \{0; 2\}$. As discussed previously, μ_f plays a crucial role when it comes to local and global identification. Finally, we consider $N = \{50; 200; 500\}$ and $T = \{5; 10; 20; 50\}$. The number of replications equals $4,000$ for each design and the factors are drawn in each replication.

Comments

Since we are interested in emphasizing different theoretical properties of the proposed linear and non-linear methods, we present results for different quantities. In particular, for the linear approach we consider the estimator itself, and results are reported in terms of mean bias, RMSE, rejection frequencies of the t-test statistic and the overidentifying restrictions (J-)test statistic (nominal size is 5% in both cases). This estimator is implemented using a single weight $q_i = 1$ with three instruments $\boldsymbol{z}_{i,t} = (y_{i,t-1}, y_{i,t-2}, y_{i,t-3})'$. Thus, the degree of overidentification is 2.

For the non-linear approach we focus on robust inference, and study the properties of the Anderson-Rubin test statistic. In this regard, this study is the first one of its kind in the panel data literature. In order to implement the procedure the set of instruments is limited to $\boldsymbol{z}_{i,t} = (y_{i,t-1}, y_{i,t-2})'$. However, in order to achieve global identification we consider three sets of weights $\boldsymbol{q}_{i,t} = (1, y_{i,t-2}, y_{i,t-3})'$. The time-invariant weight is used

for both sets of instruments, while the time-varying weights are used in pairs with the corresponding lagged values, i.e. for $z_{i,t} = y_{i,t-1}$ we set $q_{i,t} = y_{i,t-2}$, whereas for $z_{i,t} = y_{i,t-2}$ we set $q_{i,t} = y_{i,t-3}$. This specific choice corresponds to the use of two lagged values of differenced instruments in the setup of Anderson and Hsiao (1982). Thus the total number of moment conditions is 4.

The parameter μ_f plays a non-trivial role for local and global identification. In particular, for $\mu_f \neq 0$ the linear estimator is $\sqrt{NT}-$consistent and asymptotically mixed normal. Also the non-linear approach with both time-invariant and time-varying weights is globally and locally identified as long as $\mu_f \neq 0$. Thus the AR statistic will have non-trivial power towards $\alpha_A = \alpha_0 + (NT)^{-1/2}\Delta$ alternatives.

On the other hand, for $\mu_f = 0$ local and global identification might fail when the third moment of f_t is zero, as it is the case in the current setup. In particular, the linear estimator is only $\sqrt{N}-$consistent but remains asymptotically mixed normal.

Results

At first we summarize the results presented in Table S.1 for the linear estimator.

- *(Estimation)* The bias of the linear estimator developed in this paper is negligible for all combinations of N and T we consider. This is in line with our theoretical results. Furthermore, the RMSE decreases in N confirming the minimal $\sqrt{N}-$consistency of the estimator. On the other hand, the RMSE decreases in T only when $\mu_f \neq 0$, as for $\mu_f = 0$ the expected Jacobian matrix is asymptotically singular for $T \to \infty$. Both bias and RMSE are somewhat larger for $\alpha = 0.4$, than for $\alpha = 0.8$

- *(Inference)* For smaller values of T and N the t-statistic appears to be size-distorted. However, once T and (especially) N increase the rejection frequencies approach the nominal 5% level. The rejection frequencies for the $J-$statistic are close to the nominal size for almost all setups.

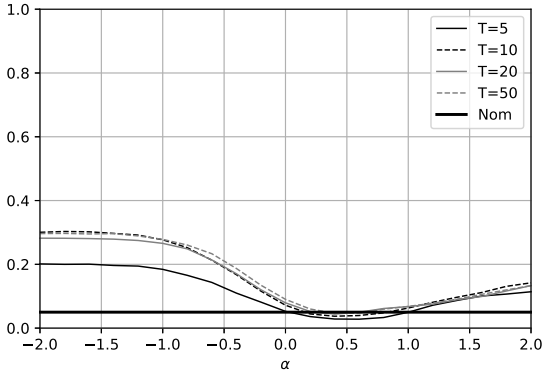
Given that for the non-linear approach we solely focus on inference, the corresponding rejection frequencies are summarized by means of Figures S.1-S.3. Below we summarize main patterns observed in these figures.

- *(Size)* The rejection frequencies under the null hypothesis are close to the nominal size of 5% in almost all cases. Only when $N = 50, T = 5, \alpha = 0.4$ the test statistic is somewhat undersized (as in the case of the linear estimator).

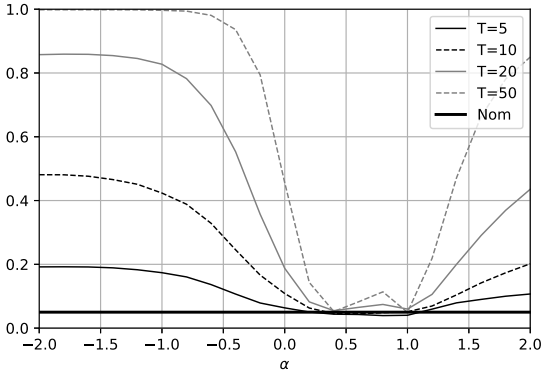
- *(The role of μ_f)* As with the linear estimator, results differ substantially, depending on the value of μ_f. In particular, while the Jacobian is non-singular for all T, including $T \to \infty$, its value is small in absolute terms when $\mu_f = 0$, thus leading to a "semi-weak instruments" problem. As a result, power is mainly with respect to $N \to \infty$, i.e. towards $\alpha_A = \alpha_0 + N^{-1/2}\Delta$ alternatives.

- *(The role of T)* For $\mu_f = 2$, the test statistic has non-trivial power over distant alternatives. In particular, for any fixed value of α and N, power is uniformly increasing in T over alternatives. This is not the case for $\mu_f = 0$.

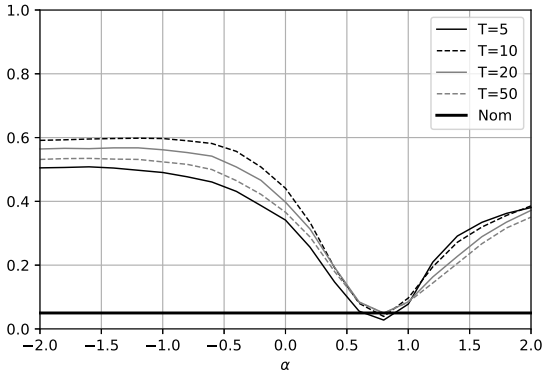
- *(Multimodality)* For $\mu_f = 0$ we can see that in most cases the power curves have a single minimum at the true value $\alpha = \alpha_0$. On the other hand, for $\mu_f = 2$ the global identification problems become more pronounced, and the power curves appear to have a second minimum at $\alpha = 1$. This is especially pronounced for $\alpha_0 = 0.4$, and for larger values of T. However, for relatively large values of N, power at $\alpha = 1$ remains above the nominal size of the test.



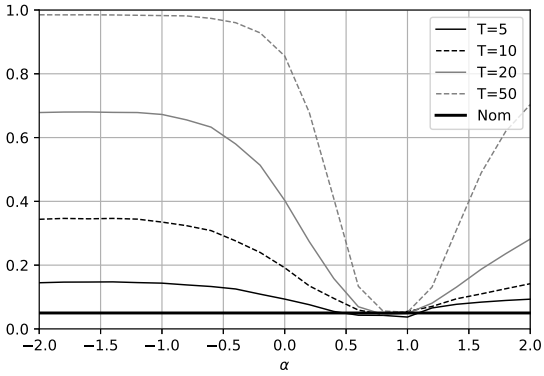
(a) $\alpha = 0.4$ and $\mu_f = 0$



(b) $\alpha = 0.4$ and $\mu_f = 2$

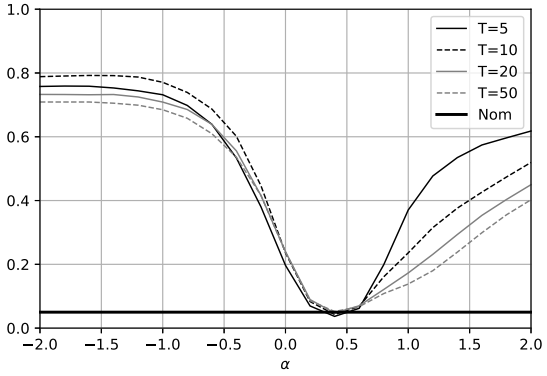


(c) $\alpha = 0.8$ and $\mu_f = 0$

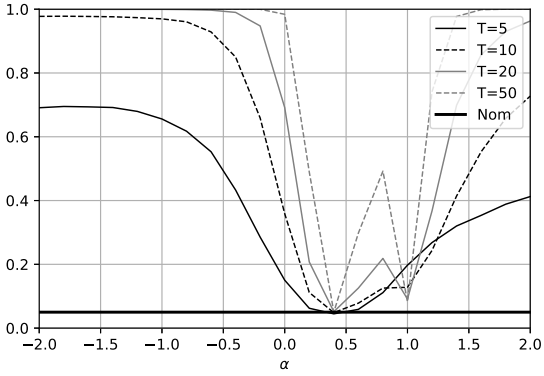


(d) $\alpha = 0.8$ and $\mu_f = 2$

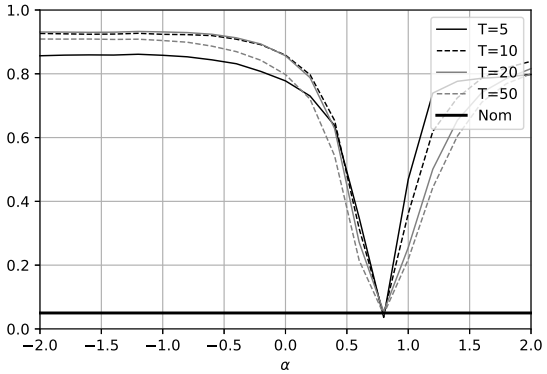
Figure S.1: Anderson-Rubin statistic rejection frequencies for $N = 50$.



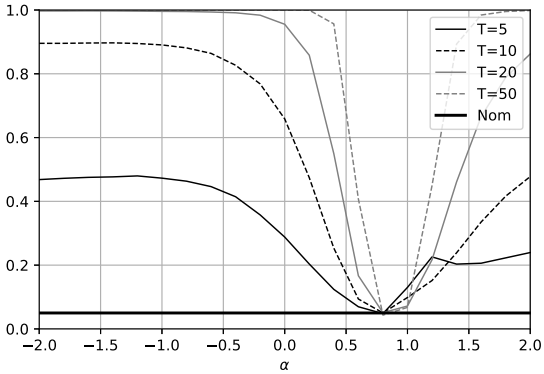
(a) $\alpha = 0.4$ and $\mu_f = 0$



(b) $\alpha = 0.4$ and $\mu_f = 2$

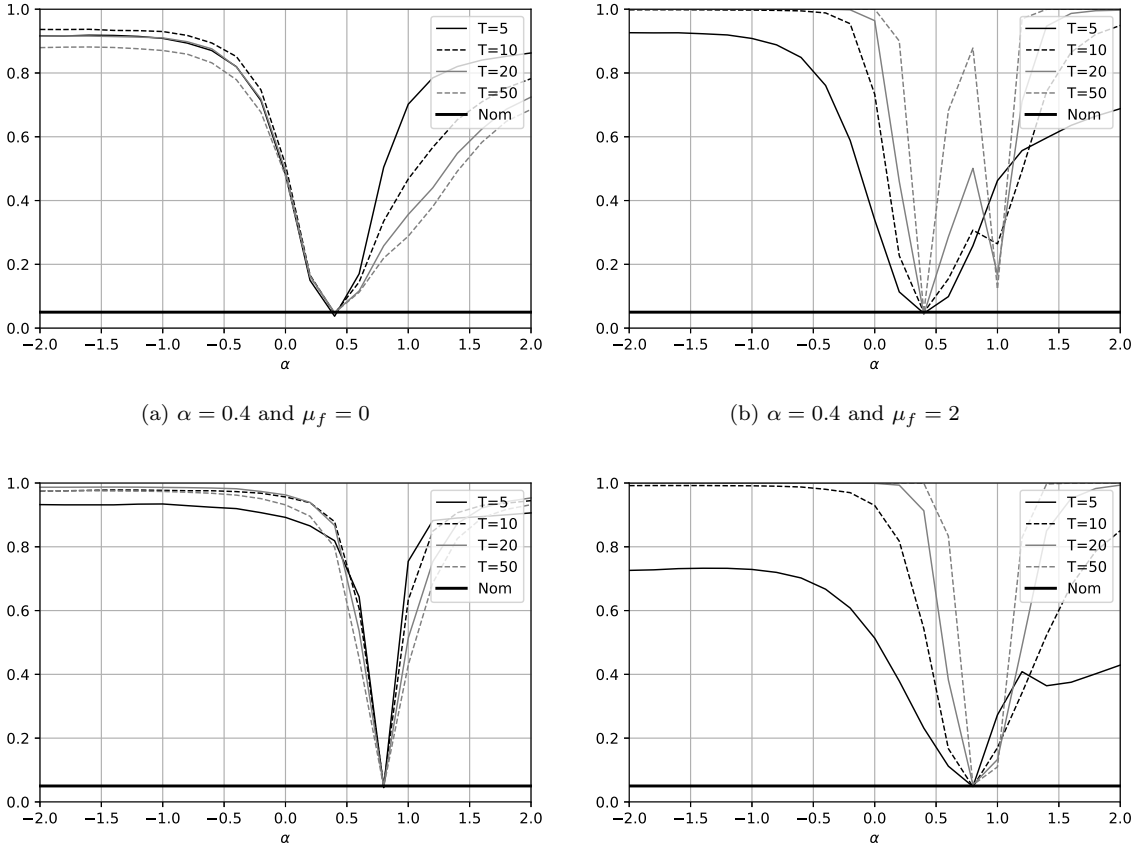


(c) $\alpha = 0.8$ and $\mu_f = 0$



(d) $\alpha = 0.8$ and $\mu_f = 2$

Figure S.2: Anderson-Rubin statistic rejection frequencies for $N = 200$.



(a) $\alpha = 0.4$ and $\mu_f = 0$

(b) $\alpha = 0.4$ and $\mu_f = 2$

(c) $\alpha = 0.8$ and $\mu_f = 0$

(d) $\alpha = 0.8$ and $\mu_f = 2$

Figure S.3: Anderson-Rubin statistic rejection frequencies for $N = 500$.

Table S.1: Linear GMM estimator with $q_i = 1$

			$N = 50$							$N = 200$							$N = 500$						
Designs			One Step GMM			Two Step GMM				One Step GMM			Two Step GMM				One Step GMM			Two Step GMM			
T	α	μ_f	Bias	RMSE	t	Bias	RMSE	t	J	Bias	RMSE	t	Bias	RMSE	t	J	Bias	RMSE	t	Bias	RMSE	t	J
5	.4	0	.002	.300	.075	-.002	.252	.116	.028	-.002	.140	.071	-.001	.115	.081	.035	.000	.100	.082	.000	.079	.089	.045
5	.4	2	.011	.158	.097	.014	.137	.126	.051	.002	.073	.080	.002	.060	.086	.049	.001	.050	.078	.001	.039	.085	.052
5	.8	0	.001	.148	.084	-.001	.126	.118	.026	-.002	.073	.078	-.002	.055	.089	.044	.000	.055	.072	-.001	.038	.084	.039
5	.8	2	.004	.065	.116	.007	.054	.142	.051	.002	.033	.082	.002	.026	.096	.048	.000	.020	.077	.001	.015	.083	.046
10	.4	0	.000	.183	.051	-.002	.167	.074	.036	.000	.093	.058	.000	.079	.060	.043	.001	.056	.056	.000	.050	.056	.051
10	.4	2	.004	.146	.065	.011	.119	.083	.041	.002	.075	.060	.003	.063	.064	.044	.001	.042	.065	.001	.035	.063	.050
10	.8	0	-.003	.145	.057	-.002	.136	.078	.034	.000	.068	.063	.000	.050	.068	.047	.000	.046	.061	.000	.032	.071	.048
10	.8	2	.003	.074	.066	.005	.051	.092	.038	.001	.039	.068	.002	.023	.071	.049	.001	.027	.062	.001	.017	.064	.049
20	.4	0	-.001	.156	.051	-.002	.141	.069	.040	-.001	.092	.057	-.001	.069	.058	.047	.001	.051	.057	.001	.045	.061	.048
20	.4	2	.000	.098	.059	.003	.085	.073	.040	.000	.050	.056	.001	.043	.055	.048	.000	.031	.060	.001	.024	.064	.046
20	.8	0	.001	.120	.044	.000	.096	.063	.033	.001	.058	.051	.000	.046	.055	.046	.000	.037	.052	.000	.027	.048	.050
20	.8	2	.002	.069	.053	.002	.042	.079	.042	.000	.053	.057	.001	.023	.066	.044	.000	.024	.054	.000	.013	.056	.048
50	.4	0	.000	.142	.042	.002	.124	.057	.043	-.001	.078	.050	.000	.067	.054	.045	.000	.047	.049	.001	.040	.053	.049
50	.4	2	.001	.048	.053	.002	.046	.070	.045	.000	.023	.052	.001	.022	.053	.046	.000	.015	.046	.000	.014	.047	.050
50	.8	0	.000	.114	.045	.001	.090	.061	.037	-.001	.058	.047	.000	.039	.051	.043	.000	.039	.048	.000	.024	.047	.048
50	.8	2	.001	.037	.059	.001	.026	.073	.042	.000	.019	.055	.000	.012	.058	.051	.000	.011	.056	.000	.008	.058	.049

Notes. Results based on $N = 4,000$ Monte Carlo draws. "Bias" corresponds to the mean bias; "RMSE" the Root Mean Squared Error; "t" correspond to the empirical rejection frequencies of the Wald test statistic with a 5% nominal level; "J" correspond to the empirical rejection frequencies of overidentifying J test statistic with a 5% nominal level.

References

Ahn, S. C., Y. H. Lee, and P. Schmidt (2013): "Panel Data Models with Multiple Time-varying Individual Effects," *Journal of Econometrics*, 174, 1–14.

Anderson, T. W. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Anderson, T. W. and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K. (2005): "Cross-Section Regression with Common Shocks," *Econometrica*, 73, 1551–1585.

Andrews, I. (2016): "Conditional Linear Combination Tests for Weakly Identified Models," *Econometrica*, 84, 2155–2182.

——— (2018): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *The Review of Economics and Statistics*, 100, 337–348.

Andrews, I. and A. Mikusheva (2016): "Conditional Inference With a Functional Nuisance Parameter," *Econometrica*, 84, 1571–1612.

Bun, M. J. G. and F. R. Kleibergen (2016): "Identification and Inference in Moments Based Analysis of Linear Dynamic Panel Data Models," UvA-Econometrics Working Paper Series.

Bun, M. J. G. and V. Sarafidis (2015): "Dynamic Panel Data Models," in *The Oxford Handbook of Panel Data*, ed. by B. H. Baltagi, Oxford: Oxford University Press, chap. 3.

Chaudhuri, S. and E. Zivot (2011): "A New Method of Projection-based Inference in GMM with Weakly Identified Nuisance Parameters," *Journal of Econometrics*, 164, 239 – 251.

Dovonon, P., A. Hall, and F. R. Kleibergen (2020): "Inference in Second-Order Identified Models," *Journal of Econometrics*, (forthcoming).

Dufour, J.-M. and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351–1365.

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Application*, Probability and Mathematical Statistics, Academic Press.

Hayakawa, K. (2009): "On the effect of mean-nonstationarity in dynamic panel data models," *Journal of Econometrics*, 153, 133–135.

Juodis, A. and V. Sarafidis (2018): "Fixed T Dynamic Panel Data Estimators with Multi-Factor Errors," *Econometric Reviews*, 37, 893–929.

——— (2020): "A Linear Estimator for Factor-Augmented Fixed-T Panels With Endogenous Regressors," *Journal of Business & Economic Statistics*, (forthcoming).

Kleibergen, F. R. (2005): "Testing Parameters in GMM without Assuming that They are Identified," *Econometrica*, 73, 1103–1123.

Kuersteiner, G. and I. R. Prucha (2013): "Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogeneity," *Journal of Econometrics*, 174, 107–126.

——— (2020): "Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity," *Econometrica*, (forthcoming).

Moreira, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Amsterdam: Elsevier, vol. 4, chap. 36, 2111–2245.

Robertson, D. and V. Sarafidis (2015): "IV Estimation of Panels with Factor Residuals," *Journal of Econometrics*, 185, 526–541.

Robertson, D., V. Sarafidis, and J. Westerlund (2018): "Unit Root Inference in Generally Trending and Cross-Correlated Dynamic Panels," *Journal of Business & Economic Statistics*, 36, 493–504.

Stock, J. H. and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.