ONLINE APPENDIX TO "SIMPLE DIFFERENCE-IN-DIFFERENCES ESTIMATION IN FIXED-T PANELS"

Nicholas Brown
Queen's University

Kyle Butts
University of Colorado Boulder

Joakim Westerlund*
Lund University
and
Deakin University

May 24, 2023

Abstract

This appendix provides (i) the proof of Theorem 1 reported in the main paper, and (ii) discussion of some of the assumptions.

1 Proof of Theorem 1

We start with part (a) of the theorem. We begin by considering the step-1 estimator of $\mathbf{f}_t$. In so doing, it is useful to denote by $\bar{\mathbf{a}}_t = (|\mathcal{I}_\infty|)^{-1} \sum_{i \in \mathcal{I}_\infty} \mathbf{a}_{i,t}$ the cross-sectional average of any vector $\mathbf{a}_{i,t}$ for the group of untreated units ($g = \infty$). In this notation, $\widehat{\mathbf{f}}_t = \bar{\mathbf{z}}_t$. Making use of this and the expression given for $\mathbf{z}_{i,t}$ in the main paper,

$$\widehat{\mathbf{f}}_t = \bar{\mathbf{z}}_t = \overline{\boldsymbol{\Lambda}}' \mathbf{f}_t + \bar{\mathbf{e}}_t \tag{A.1}$$

for the pretreatment sample $t \leq g_{\min}$. Here, $\overline{\boldsymbol{\Lambda}}$ and $\bar{\mathbf{e}}_t$ are the cross-sectional averages of $\boldsymbol{\Lambda}_i = [\boldsymbol{\alpha}_i + \boldsymbol{\lambda}_i \boldsymbol{\beta}_i, \boldsymbol{\lambda}_i]$ and $\mathbf{e}_{i,t} = [\varepsilon_{i,t} + \boldsymbol{\beta}_i' \mathbf{v}_{i,t}, \mathbf{v}_{i,t}']'$, respectively. If $m + 1 = r$, then the $r \times (m+1)$ matrix $\overline{\boldsymbol{\Lambda}}$ is square and invertible, which means that (A.1) can be rewritten as

$$\overline{\boldsymbol{\Lambda}}^{-1\prime} \widehat{\mathbf{f}}_t = \mathbf{f}_t + \overline{\boldsymbol{\Lambda}}^{-1\prime} \bar{\mathbf{e}}_t. \tag{A.2}$$

---

*Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: joakim.westerlund@nek.lu.se.

Hence, because $\|\bar{\mathbf{e}}_t\| = O_p(N^{-1/2})$ under Assumption 4, we have

$$\overline{\Lambda}^{-1\prime}\widehat{\mathbf{f}}_t = \mathbf{f}_t + O_p(N^{-1/2}) \tag{A.3}$$

and hence $\overline{\Lambda}^{-1\prime}\widehat{\mathbf{f}}_t$ is consistent for $\mathbf{f}_t$. In practice, we never observe $\overline{\Lambda}$. However, since $\boldsymbol{\alpha}_i'\mathbf{f}_t = \boldsymbol{\alpha}_i'\overline{\Lambda}^{-1\prime}\widehat{\mathbf{f}}_t + O_p(N^{-1/2})$, it is enough if we know $\widehat{\mathbf{f}}_t$, because $\overline{\Lambda}^{-1}$ is subsumed in the estimation of the coefficient of $\widehat{\mathbf{f}}_t$, which is $\mathbf{a}_i$ in our notation.

The above analysis is not possible when $m + 1 > r$ since $\overline{\Lambda}$ is no longer invertible. However, we still need something similar to (A.2), because it determines the object that is being estimated. The way we approach this issue is the same as in Westerlund et al. (2019), and others. In particular, we begin by partitioning Λ_i as $\overline{\Lambda} = [\overline{\Lambda}_r, \overline{\Lambda}_{-r}]$, where $\overline{\Lambda}_{-r}$ is $r \times (m + 1 - r)$ and $\overline{\Lambda}_r$ is $r \times r$ and full rank. Note that this partition is without loss of generality under Assumption 6. We then introduce the following $(m + 1) \times (m + 1)$ rotation matrix, which is chosen such that $\overline{\Lambda}\overline{\mathbf{H}} = [\mathbf{I}_r, \mathbf{0}_{r \times (m+1-r)}]$ and that is going to play the same role as $\overline{\Lambda}^{-1}$ under $m + 1 = r$:

$$\overline{\mathbf{H}} = \begin{bmatrix} \overline{\Lambda}_r^{-1} & -\overline{\Lambda}_r^{-1}\overline{\Lambda}_{-r} \\ \mathbf{0}_{(m+1-r) \times r} & \mathbf{I}_{m+1-r} \end{bmatrix} = [\overline{\mathbf{H}}_r, \overline{\mathbf{H}}_{-r}], \tag{A.4}$$

where $\overline{\mathbf{H}}_r = [\overline{\Lambda}_r^{-1\prime}, \mathbf{0}_{r \times (m+1-r)}]'$ is $(m + 1) \times r$, while $\overline{\mathbf{H}}_{-r} = [-\overline{\Lambda}_{-r}'\overline{\Lambda}_r^{-1\prime}, \mathbf{I}_{m+1-r}]'$ is $(m + 1) \times (m + 1 - r)$. If $m + 1 = r$, we define $\overline{\mathbf{H}} = \overline{\mathbf{H}}_r = \overline{\Lambda}_r^{-1} = \overline{\Lambda}^{-1}$. We further introduce the $(m + 1) \times (m + 1)$ matrix $\mathbf{D}_N = \text{diag}(\mathbf{I}_r, \sqrt{N}\mathbf{I}_{m+1-r})$ with $\mathbf{D}_N = \mathbf{I}_{m+1}$ if $m + 1 = r$. By pre-multiplying $\widehat{\mathbf{f}}_t$ by $\mathbf{D}_N\overline{\mathbf{H}}'$, we obtain

$$\mathbf{D}_N\overline{\mathbf{H}}'\widehat{\mathbf{f}}_t = \widehat{\mathbf{f}}_t^0 = \mathbf{D}_N\overline{\mathbf{H}}'\overline{\Lambda}'\mathbf{f}_t + \mathbf{D}_N\overline{\mathbf{H}}'\bar{\mathbf{e}}_t = \mathbf{f}_t^0 + \bar{\mathbf{e}}_t^0, \tag{A.5}$$

where $\mathbf{f}_t^0 = [\mathbf{f}_t', \mathbf{0}_{(m+1-r) \times 1}']'$ and $\bar{\mathbf{e}}_t^0 = [\bar{\mathbf{e}}_t'\overline{\mathbf{H}}_r, \sqrt{N}\bar{\mathbf{e}}_t'\overline{\mathbf{H}}_{-r}]' = [\bar{\mathbf{e}}_{r,t}^{0\prime}, \bar{\mathbf{e}}_{-r,t}^{0\prime}]'$ are both $(m + 1) \times 1$ with $\bar{\mathbf{e}}_{r,t}^0$ and $\bar{\mathbf{e}}_{-r,t}^0$ being $r \times 1$ and $(m + 1 - r) \times 1$, respectively. Hence, since $\|\bar{\mathbf{e}}_{r,t}^0\| = O_p(N^{-1/2})$ and $\|\bar{\mathbf{e}}_{-r,t}^0\| = O_p(1)$, when $m + 1 > r$ we are no longer estimating $\mathbf{f}_t$ but rather $\mathbf{f}_t^+ = [\mathbf{f}_t', \bar{\mathbf{e}}_{-r,t}^{0\prime}]'$;

$$\widehat{\mathbf{f}}_t^0 = \mathbf{f}_t^0 + \bar{\mathbf{e}}_t^0 = \begin{bmatrix} \mathbf{f}_t \\ \mathbf{0}_{(m+1-r) \times 1} \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{e}}_{r,t}^0 \\ \bar{\mathbf{e}}_{-r,t}^0 \end{bmatrix} = \mathbf{f}_t^+ + O_p(N^{-1/2}), \tag{A.6}$$

The fact that $\mathbf{f}_t$ is included in $\mathbf{f}_t^+$ suggests that asymptotically C²ED² should be able to account for the unknown factors even if $m + 1 > r$. By ensuring the existence of $\overline{\mathbf{H}}$, Assumption

6 makes this possible. However, we also note that because of the presence of $\bar{\mathbf{e}}^0_{-r,t}$, the asymptotic distribution theory will in general depend on whether $m + 1 = r$ or $m + 1 > r$.

It is useful to be able to use the above notation not only when $m + 1 > r$ but also when $m + 1 = r$. We therefore define $\widehat{\mathbf{f}}^0_t = \overline{\boldsymbol{\Lambda}}^{-1\prime}\widehat{\mathbf{f}}_t$, $\mathbf{f}^0_t = \mathbf{f}_t$ and $\bar{\mathbf{e}}^0_t = \overline{\boldsymbol{\Lambda}}^{-1\prime}\bar{\mathbf{e}}_t$ if $m + 1 = r$, so that we are back in (A.2).

Let us now consider $\widehat{\Delta}_{i,g,t}$, which, unlike $\widehat{\mathbf{f}}_t$, is computed based on treated units in post-treatment periods ($i \in \mathcal{I}_g \subset \mathcal{G}$ and $t \geq g_{\min}$). Note first that because we are considering treated units in post-treatment periods, $y_{i,t}(g) = y_{i,t}$ and $\mathbf{x}_{i,t}(g) = \mathbf{x}_{i,t}$. Further use of the definitions of $\Delta_{i,g,t}$, $y_{i,t}(\infty)$ and $\boldsymbol{\tau}_{i,g,t}$, leads to the following model for $y_{i,t}$:

$$
\begin{aligned}
y_{i,t} &= \Delta_{i,g,t} + y_{i,t}(\infty) = \Delta_{i,g,t} + \boldsymbol{\beta}'_i\mathbf{x}_{i,t}(\infty) + \boldsymbol{\alpha}'_i\mathbf{f}_t + \varepsilon_{i,t} \\
&= \Delta_{i,g,t} + \boldsymbol{\beta}'_i(\mathbf{x}_{i,t} - \boldsymbol{\tau}_{i,g,t}) + \boldsymbol{\alpha}'_i\mathbf{f}_t + \varepsilon_{i,t} = (\Delta_{i,g,t} - \boldsymbol{\beta}'_i\boldsymbol{\tau}_{i,g,t}) + \boldsymbol{\beta}'_i\mathbf{x}_{i,t} + \boldsymbol{\alpha}'_i\mathbf{f}_t + \varepsilon_{i,t} \\
&= \eta_{i,g,t} + \boldsymbol{\beta}'_i\mathbf{x}_{i,t} + \boldsymbol{\alpha}'_i\mathbf{f}_t + \varepsilon_{i,t}.
\end{aligned}
\tag{A.7}
$$

It follows that

$$
\begin{aligned}
\widehat{\Delta}_{i,g,t} &= y_{i,t} - \widehat{y}_{i,t}(\infty) \\
&= \eta_{i,g,t} + \boldsymbol{\beta}'_i\mathbf{x}_{i,t} + \boldsymbol{\alpha}'_i\mathbf{f}_t + \varepsilon_{i,t} - [\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{x}}_{i,t}(\infty) + \widehat{\mathbf{a}}'_i\widehat{\mathbf{f}}_t] \\
&= \eta_{i,g,t} + \boldsymbol{\beta}'_i\mathbf{x}_{i,t} + \boldsymbol{\alpha}'_i\mathbf{f}_t + \varepsilon_{i,t} - (\widehat{\boldsymbol{\beta}}'\mathbf{x}_{i,t} + \widehat{\mathbf{a}}'_i\widehat{\mathbf{f}}_t) + \widehat{\boldsymbol{\beta}}'[\mathbf{x}_{i,t} - \widehat{\mathbf{x}}_{i,t}(\infty)] \\
&= \eta_{i,g,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)'\mathbf{x}_{i,t} - (\widehat{\mathbf{a}}'_i\widehat{\mathbf{f}}_t - \boldsymbol{\alpha}'_i\mathbf{f}_t) + \widehat{\boldsymbol{\beta}}'[\mathbf{x}_{i,t} - \widehat{\mathbf{x}}_{i,t}(\infty)] + \varepsilon_{i,t}.
\end{aligned}
\tag{A.8}
$$

Consider $\widehat{\mathbf{a}}'_i\widehat{\mathbf{f}}_t - \boldsymbol{\alpha}'_i\mathbf{f}_t$. While the $(m + 1) \times r$ matrix $\mathbf{D}_N\overline{\mathbf{H}}'\overline{\boldsymbol{\Lambda}}'$ is not necessarily square under Assumption 6, it has full column rank. This means that we can compute its Moore–Penrose inverse, which is given by $(\mathbf{D}_N\overline{\mathbf{H}}'\overline{\boldsymbol{\Lambda}}')^+ = (\mathbf{D}_N\overline{\mathbf{H}}'\overline{\boldsymbol{\Lambda}}')' = [\mathbf{I}_r, \mathbf{0}_{r\times(m+1-r)}]$, such that $(\mathbf{D}_N\overline{\mathbf{H}}'\overline{\boldsymbol{\Lambda}}')^+\mathbf{D}_N\overline{\mathbf{H}}'\overline{\boldsymbol{\Lambda}}' = \mathbf{I}_r$. Hence, $\mathbf{D}_N\overline{\mathbf{H}}'\overline{\boldsymbol{\Lambda}}'\mathbf{f}_t = [\mathbf{f}'_t, \mathbf{0}'_{(m+1-r)\times 1}]' = \mathbf{f}^0_t$ and we also have $\mathbf{D}_N\overline{\mathbf{H}}'\widehat{\mathbf{f}}_t = \widehat{\mathbf{f}}^0_t$. Making use of this, and letting $\widehat{\mathbf{a}}^0_i = (\mathbf{D}_N\overline{\mathbf{H}}')^{-1\prime}\widehat{\mathbf{a}}_i = (\overline{\mathbf{H}}\mathbf{D}_N)^{-1}\widehat{\mathbf{a}}_i$ and $\boldsymbol{\alpha}^0_i = (\mathbf{D}_N\overline{\mathbf{H}}'\overline{\boldsymbol{\Lambda}}')^{+\prime}\boldsymbol{\alpha}_i = \mathbf{D}_N\overline{\mathbf{H}}'\overline{\boldsymbol{\Lambda}}'\boldsymbol{\alpha}_i = [\boldsymbol{\alpha}'_i, \mathbf{0}_{1\times(m+1-r)}]'$,

$$
\begin{aligned}
\widehat{\mathbf{a}}'_i\widehat{\mathbf{f}}_t - \boldsymbol{\alpha}'_i\mathbf{f}_t &= \widehat{\mathbf{a}}'_i(\mathbf{D}_N\overline{\mathbf{H}}')^{-1}\mathbf{D}_N\overline{\mathbf{H}}'\widehat{\mathbf{f}}_t - \boldsymbol{\alpha}'_i(\mathbf{D}_N\overline{\mathbf{H}}'\overline{\boldsymbol{\Lambda}}')^+\mathbf{D}_N\overline{\mathbf{H}}'\overline{\boldsymbol{\Lambda}}'\mathbf{f}_t \\
&= \widehat{\mathbf{a}}^{0\prime}_i\widehat{\mathbf{f}}^0_t - \boldsymbol{\alpha}^{0\prime}_i\mathbf{f}^0_t \\
&= \boldsymbol{\alpha}^{0\prime}_i(\widehat{\mathbf{f}}^0_t - \mathbf{f}^0_t) + (\widehat{\mathbf{a}}^0_i - \boldsymbol{\alpha}^0_i)'\widehat{\mathbf{f}}^0_t,
\end{aligned}
\tag{A.9}
$$

from which it follows that

$$\widehat{\Delta}_{i,g,t} = \eta_{i,g,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)'\mathbf{x}_{i,t} - \boldsymbol{\alpha}_i^{0\prime}(\widehat{\mathbf{f}}_t^0 - \mathbf{f}_t^0) - (\widehat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_i^0)'\widehat{\mathbf{f}}_t^0 + \widehat{\boldsymbol{\beta}}'[\mathbf{x}_{i,t} - \widehat{\mathbf{x}}_{i,t}(\infty)] + \varepsilon_{i,t}. \quad \text{(A.10)}$$

Amongst the terms appearing on the right-hand side of this last equation, the one involving $\widehat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_i^0$ requires most work. We therefore start with this. Note first that since $\widehat{\mathbf{a}}_i$ is estimated based on the pre-treatment period only, we have $y_{i,t} = y_{i,t}(\infty) = \boldsymbol{\beta}_i'\mathbf{x}_{i,t}(\infty) + \boldsymbol{\alpha}_i'\mathbf{f}_t + \varepsilon_{i,t}$ or, in terms of the stacked vector notation introduced in step 2 of the counterfactual estimation procedure outlined in the main paper, $\mathbf{y}_i = \mathbf{x}_i\boldsymbol{\beta}_i + \mathbf{f}\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i$, where $\mathbf{y}_i$, $\mathbf{x}_i$, $\mathbf{f}$ and $\boldsymbol{\varepsilon}_i$ are all $(g_{\min} - 1)$-rowed. By using this and $\overline{\boldsymbol{\Lambda}}\mathbf{H}_r = \mathbf{I}_r$, we get

$$\mathbf{y}_i = \mathbf{x}_i\boldsymbol{\beta}_i + \widehat{\mathbf{f}}\overline{\mathbf{H}}_r\boldsymbol{\alpha}_i - (\widehat{\mathbf{f}} - \mathbf{f}\overline{\boldsymbol{\Lambda}})\overline{\mathbf{H}}_r\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i = \mathbf{x}_i\boldsymbol{\beta}_i + \widehat{\mathbf{f}}\overline{\mathbf{H}}_r\boldsymbol{\alpha}_i - \overline{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i. \quad \text{(A.11)}$$

We also note that $\mathbf{a}_i$ in step 2 can be expressed in terms of $\overline{\mathbf{H}}_r$ and $\boldsymbol{\alpha}_i$ as $\mathbf{a}_i = \overline{\mathbf{H}}_r\boldsymbol{\alpha}_i$. By inserting this and (A.11) into the expression given for $\widehat{\mathbf{a}}_i$ in step 2,

$$\begin{aligned}
\widehat{\mathbf{a}}_i &= (\widehat{\mathbf{f}}'\widehat{\mathbf{f}})^{-1}\widehat{\mathbf{f}}'(\mathbf{y}_i - \mathbf{x}_i\widehat{\boldsymbol{\beta}}) \\
&= (\widehat{\mathbf{f}}'\widehat{\mathbf{f}})^{-1}\widehat{\mathbf{f}}'(\mathbf{x}_i\boldsymbol{\beta}_i + \widehat{\mathbf{f}}\mathbf{a}_i - \overline{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i - \mathbf{x}_i\widehat{\boldsymbol{\beta}}) \\
&= \mathbf{a}_i + (\widehat{\mathbf{f}}'\widehat{\mathbf{f}})^{-1}\widehat{\mathbf{f}}'[-\mathbf{x}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) - \overline{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i],
\end{aligned} \quad \text{(A.12)}$$

implying

$$\begin{aligned}
\widehat{\mathbf{a}}_i^0 &= (\overline{\mathbf{H}}\mathbf{D}_N)^{-1}\widehat{\mathbf{a}}_i \\
&= (\overline{\mathbf{H}}\mathbf{D}_N)^{-1}\mathbf{a}_i + (\overline{\mathbf{H}}\mathbf{D}_N)^{-1}(\widehat{\mathbf{f}}'\widehat{\mathbf{f}})^{-1}\widehat{\mathbf{f}}'[-\mathbf{x}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) - \overline{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i] \\
&= (\overline{\mathbf{H}}\mathbf{D}_N)^{-1}\mathbf{a}_i + (\mathbf{D}_N\overline{\mathbf{H}}'\widehat{\mathbf{f}}'\widehat{\mathbf{f}}\overline{\mathbf{H}}\mathbf{D}_N)^{-1}\mathbf{D}_N\overline{\mathbf{H}}'\widehat{\mathbf{f}}'[-\mathbf{x}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) - \overline{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i] \\
&= (\overline{\mathbf{H}}\mathbf{D}_N)^{-1}\mathbf{a}_i + (\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}^{0\prime}[-\mathbf{x}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) - \overline{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i]
\end{aligned} \quad \text{(A.13)}$$

where $\widehat{\mathbf{f}}^0 = [\widehat{\mathbf{f}}_1^0, ..., \widehat{\mathbf{f}}_{g_{\min}-1}^0]' = \widehat{\mathbf{f}}\overline{\mathbf{H}}\mathbf{D}_N$ is $(g_{\min} - 1) \times (m + 1)$. Consider the first term on the right-hand side. A direct calculation using the rules for the inverse of a partitioned matrix (see, for example, Abadir and Magnus (2005), Exercise 5.16) reveals that

$$(\mathbf{D}_N\overline{\mathbf{H}})^{-1} = \begin{bmatrix} \overline{\boldsymbol{\Lambda}}_r & \overline{\boldsymbol{\Lambda}}_{-r} \\ \mathbf{0}_{(m+1-r)\times r} & N^{-1/2}\mathbf{I}_{m+1-r} \end{bmatrix}, \quad \text{(A.14)}$$

so that

$$(\overline{\mathbf{H}}\mathbf{D}_N)^{-1}\overline{\mathbf{H}}_r = \begin{bmatrix} \overline{\boldsymbol{\Lambda}}_r & \overline{\boldsymbol{\Lambda}}_{-r} \\ \mathbf{0}_{(m+1-r)\times r} & N^{-1/2}\mathbf{I}_{m+1-r} \end{bmatrix} \begin{bmatrix} \overline{\boldsymbol{\Lambda}}_r^{-1} \\ \mathbf{0}_{(m+1-r)\times r} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_r \\ \mathbf{0}_{(m+1-r)\times r} \end{bmatrix}. \quad (A.15)$$

This implies

$$(\overline{\mathbf{H}}\mathbf{D}_N)^{-1}\mathbf{a}_i = \begin{bmatrix} \boldsymbol{\alpha}_i \\ \mathbf{0}_{(m+1-r)\times 1} \end{bmatrix} = \boldsymbol{\alpha}_i^0, \quad (A.16)$$

leading to the following expression for $\widehat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_i^0$:

$$\widehat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_i^0 = (\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}^{0\prime}[-\mathbf{x}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) - \overline{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i]. \quad (A.17)$$

We similarly have

$$\widehat{\mathbf{x}}_{i,t}(\infty) = \widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{f}}_t = \mathbf{x}_i'\widehat{\mathbf{f}}(\widehat{\mathbf{f}}'\widehat{\mathbf{f}})^{-1}\widehat{\mathbf{f}}_t = \mathbf{x}_i'\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0, \quad (A.18)$$

from which it follows that

$$\widehat{\boldsymbol{\beta}}'[\mathbf{x}_{i,t} - \widehat{\mathbf{x}}_{i,t}(\infty)] = \widehat{\boldsymbol{\beta}}'[\mathbf{x}_{i,t} - \mathbf{x}_i'\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0]. \quad (A.19)$$

By inserting the above expressions into the one given earlier for $\widehat{\Delta}_{i,g,t}$, we get

$$\begin{aligned}
\widehat{\Delta}_{i,g,t} &= \eta_{i,g,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)'\mathbf{x}_{i,t} - \boldsymbol{\alpha}_i^{0\prime}(\widehat{\mathbf{f}}_t^0 - \mathbf{f}_t^0) - (\widehat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_i^0)'\widehat{\mathbf{f}}_t^0 + \widehat{\boldsymbol{\beta}}'[\mathbf{x}_{i,t} - \widehat{\mathbf{x}}_{i,t}(\infty)] + \varepsilon_{i,t} \\
&= \eta_{i,g,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)'\mathbf{x}_{i,t} - \boldsymbol{\alpha}_i'\overline{\mathbf{e}}_{r,t}^0 - [-\mathbf{x}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) - \overline{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i]'\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0 \\
&\quad + \widehat{\boldsymbol{\beta}}'[\mathbf{x}_{i,t} - \mathbf{x}_i'\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0] + \varepsilon_{i,t} \\
&= \eta_{i,g,t} + \boldsymbol{\beta}_i'\mathbf{x}_{i,t} - \boldsymbol{\alpha}_i'\overline{\mathbf{e}}_{r,t}^0 + \varepsilon_{i,t} - (\mathbf{x}_i\boldsymbol{\beta}_i - \overline{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i)'\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0.
\end{aligned} \quad (A.20)$$

Further use of $\widehat{\mathbf{f}} = \widehat{\mathbf{f}}^0\mathbf{D}_N^{-1}\overline{\mathbf{H}}^{-1}$ gives

$$\mathbf{x}_i = \mathbf{f}\boldsymbol{\lambda}_i + \mathbf{v}_i = \widehat{\mathbf{f}}\overline{\mathbf{H}}_r\boldsymbol{\lambda}_i - (\widehat{\mathbf{f}} - \mathbf{f}\overline{\boldsymbol{\Lambda}})\overline{\mathbf{H}}_r\boldsymbol{\lambda}_i + \mathbf{v}_i = \widehat{\mathbf{f}}^0\mathbf{D}_N^{-1}\overline{\mathbf{H}}^{-1}\overline{\mathbf{H}}_r\boldsymbol{\lambda}_i - \overline{\mathbf{e}}_r^0\boldsymbol{\lambda}_i + \mathbf{v}_i, \quad (A.21)$$

for $t < g_{\min}$. If, on the other hand, $t \geq g_{\min}$, then

$$\mathbf{x}_{i,t} = \boldsymbol{\tau}_{i,g,t} + \boldsymbol{\lambda}_i'\mathbf{f}_t + \mathbf{v}_{i,t} = \boldsymbol{\tau}_{i,g,t} + \boldsymbol{\lambda}_i'\overline{\mathbf{H}}_r'\overline{\mathbf{H}}^{-1\prime}\mathbf{D}_N^{-1}\widehat{\mathbf{f}}_t^0 - \boldsymbol{\lambda}_i'\overline{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t}. \quad (A.22)$$

These two last results imply

$$\begin{aligned}
&\mathbf{x}_{i,t} - \mathbf{x}_i'\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0 \\
&= \boldsymbol{\tau}_{i,g,t} + \boldsymbol{\lambda}_i'\overline{\mathbf{H}}_r'\overline{\mathbf{H}}^{-1\prime}\mathbf{D}_N^{-1}\widehat{\mathbf{f}}_t^0 - \boldsymbol{\lambda}_i'\overline{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t} - (\widehat{\mathbf{f}}^0\mathbf{D}_N^{-1}\overline{\mathbf{H}}^{-1}\overline{\mathbf{H}}_r\boldsymbol{\lambda}_i - \overline{\mathbf{e}}_r^0\boldsymbol{\lambda}_i + \mathbf{v}_i)'\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0 \\
&= \boldsymbol{\tau}_{i,g,t} - \boldsymbol{\lambda}_i'\overline{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t} - (-\overline{\mathbf{e}}_r^0\boldsymbol{\lambda}_i + \mathbf{v}_i)'\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0,
\end{aligned} \quad (A.23)$$

and so we arrive at the following expression for $\widehat{\Delta}_{i,g,t}$:

$$
\begin{aligned}
\widehat{\Delta}_{i,g,t} &= \eta_{i,g,t} + \boldsymbol{\beta}_i'(\boldsymbol{\tau}_{i,g,t} - \boldsymbol{\lambda}_i'\bar{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t}) - \boldsymbol{\alpha}_i'\bar{\mathbf{e}}_{r,t}^0 + \varepsilon_{i,t} \\
&\quad - [(-\bar{\mathbf{e}}_r^0\boldsymbol{\lambda}_i + \mathbf{v}_i)\boldsymbol{\beta}_i - \bar{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \varepsilon_i]'\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0 \\
&= \Delta_{i,g,t} - (\boldsymbol{\lambda}_i\boldsymbol{\beta}_i + \boldsymbol{\alpha}_i)'\bar{\mathbf{e}}_{r,t}^0 + \boldsymbol{\beta}_i'\mathbf{v}_{i,t} + \varepsilon_{i,t} \\
&\quad - [-\bar{\mathbf{e}}_r^0(\boldsymbol{\lambda}_i\boldsymbol{\beta}_i + \boldsymbol{\alpha}_i) + \mathbf{v}_i\boldsymbol{\beta}_i + \varepsilon_i]'\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0.
\end{aligned}
\tag{A.24}
$$

where $\Delta_{i,g,t} = \eta_{i,g,t} + \boldsymbol{\beta}_i'\boldsymbol{\tau}_{i,g,t}$ as in the main paper.

The above expression for $\widehat{\Delta}_{i,g,t}$ is the cleanest possible without exploiting the fact that N is large. Hence, in what remains we are going to let $N \to \infty$. We begin by considering $\widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0$. Define $\mathbf{f}^+ = [\mathbf{f}_1^+, ... \mathbf{f}_{g_{\min}-1}^+]' = [\mathbf{f}, \bar{\mathbf{e}}_{-r}^0]$, a $(g_{\min} - 1) \times (m + 1)$ matrix. We have already shown that $\widehat{\mathbf{f}}^0 = \mathbf{f}^+ + O_p(N^{-1/2})$. By using this and the results provided in the proof of Lemma A.1 in Westerlund et al. (2019), we have that $\|\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0 - \mathbf{f}^{+\prime}\mathbf{f}^+\| = O_p(N^{-1/2})$ and, more importantly,

$$
\|(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1} - (\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\| = O_p(N^{-1/2}),
\tag{A.25}
$$

where

$$
\mathbf{f}^{+\prime}\mathbf{f}^+ = \begin{bmatrix} \mathbf{f}'\mathbf{f} & \mathbf{f}'\bar{\mathbf{e}}_{-r}^0 \\ \bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{f} & \bar{\mathbf{e}}_{-r}^{0\prime}\bar{\mathbf{e}}_{-r}^0 \end{bmatrix},
\tag{A.26}
$$

$$
(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1} = \begin{bmatrix} (\mathbf{f}'\mathbf{f})^{-1} + (\mathbf{f}'\mathbf{f})^{-1}\mathbf{f}'\bar{\mathbf{e}}_{-r}^0(\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M}_{\mathbf{f}}\bar{\mathbf{e}}_{-r}^0)^{-1}\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{f}(\mathbf{f}'\mathbf{f})^{-1} \\ -(\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M}_{\mathbf{f}}\bar{\mathbf{e}}_{-r}^0)^{-1}\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{f}(\mathbf{f}'\mathbf{f})^{-1} \\ \quad -(\mathbf{f}'\mathbf{f})^{-1}\mathbf{f}'\bar{\mathbf{e}}_{-r}^0(\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M}_{\mathbf{f}}\bar{\mathbf{e}}_{-r}^0)^{-1} \\ \quad (\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M}_{\mathbf{f}}\bar{\mathbf{e}}_{-r}^0)^{-1} \end{bmatrix}.
\tag{A.27}
$$

The expression for $(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}$ is again obtained by using the rules for the inverse of a partitioned matrix. The fact that $\|(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1} - (\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\| = O_p(N^{-1/2})$ together with $\widehat{\mathbf{f}}^0 = \mathbf{f}^+ + O_p(N^{-1/2})$ imply that

$$
\begin{aligned}
\widehat{\mathbf{f}}_t^{0\prime}(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}^{0\prime} &= \widehat{\mathbf{f}}_t^{0\prime}[(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1} - (\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}]\widehat{\mathbf{f}}^{0\prime} + \widehat{\mathbf{f}}_t^{0\prime}(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\widehat{\mathbf{f}}^{0\prime} \\
&= \widehat{\mathbf{f}}_t^{0\prime}(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\widehat{\mathbf{f}}^{0\prime} + O_p(N^{-1/2}) \\
&= \mathbf{f}_t^{+\prime}(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\mathbf{f}^{+\prime} + O_p(N^{-1/2}).
\end{aligned}
\tag{A.28}
$$

where, defining $\mathbf{M_f}$ analogously to $\mathbf{M_{\widehat{f}}}$,

$$\mathbf{f}_t^{+\prime}(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\mathbf{f}^{+\prime}$$

$$= [\mathbf{f}_t', \bar{\mathbf{e}}_{-r,t}^{0\prime}]\left[\begin{array}{c} (\mathbf{f}'\mathbf{f})^{-1} + (\mathbf{f}'\mathbf{f})^{-1}\mathbf{f}'\bar{\mathbf{e}}_{-r}^0(\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M_f}\bar{\mathbf{e}}_{-r}^0)^{-1}\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{f}(\mathbf{f}'\mathbf{f})^{-1} \\ -(\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M_f}\bar{\mathbf{e}}_{-r}^0)^{-1}\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{f}(\mathbf{f}'\mathbf{f})^{-1} \end{array}\right.$$

$$\left.\begin{array}{c} -(\mathbf{f}'\mathbf{f})^{-1}\mathbf{f}'\bar{\mathbf{e}}_{-r}^0(\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M_f}\bar{\mathbf{e}}_{-r}^0)^{-1} \\ (\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M_f}\bar{\mathbf{e}}_{-r}^0)^{-1} \end{array}\right]\left[\begin{array}{c} \mathbf{f}' \\ \bar{\mathbf{e}}_{-r}^{0\prime} \end{array}\right]$$

$$= \mathbf{f}_t'(\mathbf{f}'\mathbf{f})^{-1}\mathbf{f}'[\mathbf{I}_{g_{\min}-1} - \bar{\mathbf{e}}_{-r}^0(\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M_f}\bar{\mathbf{e}}_{-r}^0)^{-1}\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M_f}] + \bar{\mathbf{e}}_{-r,t}^{0\prime}(\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M_f}\bar{\mathbf{e}}_{-r}^0)^{-1}\bar{\mathbf{e}}_{-r}^{0\prime}\mathbf{M_f}. \qquad (A.29)$$

The fact that $\|\widehat{\mathbf{f}}_t^{0\prime}(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}^{0\prime} - \mathbf{f}_t^{+\prime}(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\mathbf{f}^{+\prime}\| = O_p(N^{-1/2})$ implies

$$\widehat{\Delta}_{i,g,t} = \Delta_{i,g,t} - (\lambda_i\beta_i + \alpha_i)'\bar{\mathbf{e}}_{r,t}^0 + \beta_i'\mathbf{v}_{i,t} + \varepsilon_{i,t}$$

$$- [-\bar{\mathbf{e}}_r^0(\lambda_i\beta_i + \alpha_i) + \mathbf{v}_i\beta_i + \varepsilon_i]'\mathbf{f}^+(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\mathbf{f}_t^+ + O_p(N^{-1/2})$$

$$= \Delta_{i,g,t} - (\lambda_i\beta_i + \alpha_i)'\bar{\mathbf{e}}_{r,t}^{0*} + \beta_i'\mathbf{v}_{i,t}^* + \varepsilon_{i,t}^* + O_p(N^{-1/2}), \qquad (A.30)$$

where

$$\mathbf{a}_{i,t}^* = \mathbf{a}_{i,t} - \mathbf{a}_i'\mathbf{f}^+(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\mathbf{f}_t^+ = \mathbf{a}_{i,t} - \sum_{s=1}^{g_{\min}-1}\mathbf{a}_{i,s}\mathbf{f}_s^{+\prime}(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\mathbf{f}_t^+ \qquad (A.31)$$

for any vector $\mathbf{a}_{i,t}$ with $(g_{\min} - 1)$-rowed stack $\mathbf{a}_i = [\mathbf{a}_{i,1}, ..., \mathbf{a}_{i,g_{\min}-1}]'$. In words, $\mathbf{a}_{i,t}^*$ is the limiting "defactored" version of $\mathbf{a}_{i,t}$.

We now make use of the above expression for $\widehat{\Delta}_{i,g,t}$ to evaluate $\widehat{\Delta}_{g,t}$. In so doing, it is important to note that the order of the reminder incurred when replacing $\widehat{\mathbf{f}}_t^{0\prime}(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}^{0\prime}$ with $\mathbf{f}_t^{+\prime}(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\mathbf{f}^{+\prime}$ is the same even after averaging over group g and multiplying by $\sqrt{|\mathcal{I}_g|}$. In order to appreciate this, we make use of the fact that $\|\sqrt{|\mathcal{I}_g|}\bar{\mathbf{e}}_r^0\| = O_p(1)$, and since $\mathbf{v}_i$ and β_i are independent with $\mathbf{v}_i$ mean zero and independent also across i, we also have $\|(|\mathcal{I}_g|)^{-1/2}\sum_{i\in\mathcal{I}_g}\mathbf{v}_i\beta_i\| = O_p(1)$. It follows that

$$\left\|\frac{1}{\sqrt{|\mathcal{I}_g|}}\sum_{i\in\mathcal{I}_g}[-\bar{\mathbf{e}}_r^0(\lambda_i\beta_i + \alpha_i) + \mathbf{v}_i\beta_i + \varepsilon_i]\right\|$$

$$\leq \|\sqrt{|\mathcal{I}_g|}\bar{\mathbf{e}}_r^0\|\left\|\frac{1}{|\mathcal{I}_g|}\sum_{i\in\mathcal{I}_g}(\lambda_i\beta_i + \alpha_i)\right\| + \left\|\frac{1}{\sqrt{|\mathcal{I}_g|}}\sum_{i\in\mathcal{I}_g}\mathbf{v}_i\beta_i\right\| + \left\|\frac{1}{\sqrt{|\mathcal{I}_g|}}\sum_{i\in\mathcal{I}_g}\varepsilon_i\right\| = O_p(1).$$

$$(A.32)$$

We can therefore show that

$$\left\| \frac{1}{\sqrt{|\mathcal{I}_g|}} \sum_{i \in \mathcal{I}_g} [-\bar{\mathbf{e}}_r^0(\lambda_i \boldsymbol{\beta}_i + \boldsymbol{\alpha}_i) + \mathbf{v}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i]'[\mathbf{f}^+(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\mathbf{f}_t^+ - \widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0] \right\|$$

$$\leq \left\| \frac{1}{\sqrt{|\mathcal{I}_g|}} \sum_{i \in \mathcal{I}_g} [-\bar{\mathbf{e}}_r^0(\lambda_i \boldsymbol{\beta}_i + \boldsymbol{\alpha}_i) + \mathbf{v}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i] \right\| \|\mathbf{f}^+(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\mathbf{f}_t^+ - \widehat{\mathbf{f}}^0(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}_t^0\|$$

$$= O_p(N^{-1/2}), \tag{A.33}$$

which means that the reminder incurred when replacing $\widehat{\mathbf{f}}_t^{0\prime}(\widehat{\mathbf{f}}^{0\prime}\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}^{0\prime}$ with $\mathbf{f}_t^{+\prime}(\mathbf{f}^{+\prime}\mathbf{f}^+)^{-1}\mathbf{f}^{+\prime}$ is $O_p(N^{-1/2})$ after averaging over group g and multiplying by $\sqrt{|\mathcal{I}_g|}$.

For $\Delta_{i,g,t}$, we make use of the fact that $\Delta_{i,g,t} = \Delta_{g,t} + v_{i,t}$ for $i \in \mathcal{I}_g$ by Assumption 3, giving

$$\sqrt{|\mathcal{I}_g|}(\widehat{\Delta}_{g,t} - \Delta_{g,t}) = \frac{1}{\sqrt{|\mathcal{I}_g|}} \sum_{i \in \mathcal{I}_g} (\widehat{\Delta}_{i,g,t} - \Delta_{g,t})$$

$$= \frac{1}{\sqrt{|\mathcal{I}_g|}} \sum_{i \in \mathcal{I}_g} (\widehat{\Delta}_{i,g,t} - \Delta_{i,g,t} + v_{i,t})$$

$$= \frac{1}{\sqrt{|\mathcal{I}_g|}} \sum_{i \in \mathcal{I}_g} [v_{i,t} - (\lambda_i \boldsymbol{\beta}_i + \boldsymbol{\alpha}_i)'\bar{\mathbf{e}}_{r,t}^{0*} + \boldsymbol{\beta}_i'\mathbf{v}_{i,t}^* + \varepsilon_{i,t}^*] + O_p(N^{-1/2}). \tag{A.34}$$

Moreover, $|\mathcal{I}_g|/N \to_p \tau_g \in (0,1)$ by Assumption 2. Hence, if we in addition define $\mathbf{a}_g = \text{plim}_{N\to\infty}(|\mathcal{I}_g|)^{-1} \sum_{i \in \mathcal{I}_g}(\lambda_i \boldsymbol{\beta}_i + \boldsymbol{\alpha}_i)$, the above expression for $\sqrt{|\mathcal{I}_g|}(\widehat{\Delta}_{g,t} - \Delta_{g,t})$ becomes

$$\sqrt{|\mathcal{I}_g|}(\widehat{\Delta}_{g,t} - \Delta_{g,t})$$

$$= \frac{1}{\sqrt{|\mathcal{I}_g|}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \boldsymbol{\beta}_i'\mathbf{v}_{i,t}^* + \varepsilon_{i,t}^*) - \sqrt{\frac{|\mathcal{I}_g|}{N}} \frac{1}{|\mathcal{I}_g|} \sum_{i \in \mathcal{I}_g} (\lambda_i \boldsymbol{\beta}_i + \boldsymbol{\alpha}_i)'\sqrt{N}\bar{\mathbf{e}}_{r,t}^{0*} + O_p(N^{-1/2})$$

$$= \frac{1}{\sqrt{|\mathcal{I}_g|}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \boldsymbol{\beta}_i'\mathbf{v}_{i,t}^* + \varepsilon_{i,t}^*) - \sqrt{\tau_g}\mathbf{a}_g'\sqrt{N}\bar{\mathbf{e}}_{r,t}^{0*} + o_p(1). \tag{A.35}$$

All the terms on the right-hand side of the above equation are mean zero and independent across i (conditionally on $\mathbf{f}$). They are therefore asymptotically normal by a central limit law for independent variables. However, they are not uncorrelated with each other, which complicates the calculation of the asymptotic variance. Let us therefore define $\sigma^2(\widehat{\Delta}_{g,t}) = \text{var}(\sqrt{|\mathcal{I}_g|}(\widehat{\Delta}_{g,t} - \Delta_{g,t})|\mathcal{C})$, where $\mathcal{C}$ is the sigma-field generated by $\mathbf{f}$. The asymptotic distri-

bution of $\sqrt{|\mathcal{I}_g|}(\widehat{\Delta}_{g,t} - \Delta_{g,t})$ as $N \to \infty$ can now be stated in the following way:

$$\sqrt{|\mathcal{I}_g|}(\widehat{\Delta}_{g,t} - \Delta_{g,t}) \to_d MN(0, \sigma^2(\widehat{\Delta}_{g,t})), \tag{A.36}$$

where $MN(\cdot, \cdot)$ signifies a mixed normal distribution that is normal conditionally on $\mathcal{C}$. This means that the conditional distribution of $\sqrt{|\mathcal{I}_g|}(\widehat{\Delta}_{g,t} - \Delta_{g,t})/\sigma(\widehat{\Delta}_{g,t})$ is also the unconditional distribution. Hence,

$$\frac{\sqrt{|\mathcal{I}_g|}(\widehat{\Delta}_{g,t} - \Delta_{g,t})}{\sigma(\widehat{\Delta}_{g,t})} \to_d N(0,1), \tag{A.37}$$

as required for part (a).

It remains to prove (b) and the consistency of $\widehat{\sigma}^2(\widehat{\Delta}_{g,t})$. From before,

$$\widehat{\Delta}_{i,g,t} = \Delta_{g,t} + v_{i,t} - (\lambda_i \beta_i + \alpha_i)' \overline{\mathbf{e}}_{r,t}^{0*} + \beta_i' \mathbf{v}_{i,t}^* + \varepsilon_{i,t}^* + O_p(N^{-1/2}), \tag{A.38}$$

$$\frac{1}{|\mathcal{I}_g|} \sum_{i \in \mathcal{I}_g} \widehat{\Delta}_{i,g,t} = \Delta_{g,t} + \frac{1}{|\mathcal{I}_g|} \sum_{i \in \mathcal{I}_g} [v_{i,t} - (\lambda_i \beta_i + \alpha_i)' \overline{\mathbf{e}}_{r,t}^{0*} + \beta_i' \mathbf{v}_{i,t}^* + \varepsilon_{i,t}^*] + O_p(N^{-1/2}). \tag{A.39}$$

It follows that if we let $z_{i,t} = v_{i,t} - (\lambda_i \beta_i + \alpha_i)' \overline{\mathbf{e}}_{r,t}^{0*} + \beta_i' \mathbf{v}_{i,t}^* + \varepsilon_{i,t}^*$, then

$$\widehat{\Delta}_{i,g,t} - \frac{1}{|\mathcal{I}_g|} \sum_{j \in \mathcal{I}_g} \widehat{\Delta}_{j,g,t} = z_{i,t} - \frac{1}{|\mathcal{I}_g|} \sum_{j \in \mathcal{I}_g} z_{j,t} + O_p(N^{-1/2}). \tag{A.40}$$

Hence, since $z_{i,t}$ is again independent across i, by a law of large numbers for independent variables,

$$\begin{aligned}
\widehat{\sigma}^2(\widehat{\Delta}_{g,t}) &= \frac{1}{|\mathcal{I}_g| - 1} \sum_{i \in \mathcal{I}_g} \left(\widehat{\Delta}_{i,g,t} - \frac{1}{|\mathcal{I}_g|} \sum_{j \in \mathcal{I}_g} \widehat{\Delta}_{j,g,t} \right)^2 \\
&= \frac{1}{|\mathcal{I}_g| - 1} \sum_{i \in \mathcal{I}_g} \left(z_{i,t} - \frac{1}{|\mathcal{I}_g|} \sum_{j \in \mathcal{I}_g} z_{j,t} \right)^2 + O_p(N^{-1/2}) \to_p \sigma^2(\widehat{\Delta}_{g,t})
\end{aligned} \tag{A.41}$$

as $N \to \infty$ (see Pesaran, 2006, page 985, for a similar argument). This establishes part (b) and hence the proof of the theorem is complete.

2 Discussion of some of the assumptions

The results reported in the main paper assume that the covariates admits to a common factor representation, which is not needed in pricipal components-based studies such as that of

Chan and Kwok (2022). In this section, we show that the direct effect is estimable even if this assumption fails, although in that case we can no longer identify the overall and indirect ATTs.

Our starting point here is Brown et al. (2022), who consider the same CCE approach as in Pesaran (2006) but under a different set of assumptions. In particular, instead of requiring that $\mathbf{x}_{i,t}$ has factor structure, they assume that $\mathbf{f}_t$ satisfies

$$\mathbf{f}_t = \mathbf{B}'\mathbf{\Psi}_t \tag{B.42}$$

where $\mathbf{\Psi}_t = \mathbb{E}(\mathbf{z}_{i,t})$ is constant in i and $\mathbf{B}$ is an arbitrary $(m+1) \times r$ matrix of constants. Unlike the factors-in-covariates condition, (B.42) is not testable. However, if it holds, $\widehat{\mathbf{f}}_t$ can be used to estimate an arbitrary number of factors. In order to illustrate this last point, note that if (B.42) holds for the untreated potential outcomes,

$$
\begin{aligned}
y_{i,t}(\infty) &= \boldsymbol{\beta}'_i \mathbf{x}_{i,t}(\infty) + \boldsymbol{\alpha}'_i \mathbf{f}_t + \epsilon_{i,t} \\
&= \boldsymbol{\beta}'_i \mathbf{x}_{i,t}(\infty) + \mathbf{a}'_i \mathbb{E}(\mathbf{z}_{i,t}|g_i = \infty) + \epsilon_{i,t} \\
&= \boldsymbol{\beta}'_i \mathbf{x}_{i,t}(\infty) + \mathbf{a}'_i \widehat{\mathbf{f}}_t + \mathbf{a}'_i [\mathbb{E}(\mathbf{z}_{i,t}|g_i = \infty) - \widehat{\mathbf{f}}_t] + \epsilon_{i,t}
\end{aligned}
\tag{B.43}
$$

where $\mathbf{a}_i = \mathbf{B}\boldsymbol{\alpha}_i$ and $\mathbb{E}(\mathbf{z}_{i,t}|g_i = \infty) - \widehat{\mathbf{f}}_t$ is negligible, as $\bar{\mathbf{z}}_t \rightarrow_p \mathbb{E}(\mathbf{z}_{i,t})$ as $N \rightarrow \infty$ under standard regulatory conditions.

The main advantage of (B.42) is that it leaves the covariates essentially unrestricted. However, because we no longer have a model for the untreated potential covariates, we cannot estimate $\mathbf{x}_{i,t}(\infty)$ in step 3 of the counterfactual estimation procedure. This has two implications; (i) we are unable to identify the effect of the treatment on $\mathbf{x}_{i,t}$, and (ii) we have to use $\mathbf{x}_{i,t}$ as opposed to $\widehat{\mathbf{x}}_{i,t}(\infty)$ when computing $\widehat{y}_{i,t}(\infty)$ in step 4. As a result, similarly to Chan and Kwok (2022), we can only identify the direct ATT. Hence, while we can relax factors-in-covariates condition, this has a price in terms of the estimable ATTs.

References

ABADIR, K. M. AND J. R. MAGNUS (2005): *Matrix algebra*, vol. 1, Cambridge University Press.

BROWN, N. L., P. SCHMIDT, AND J. M. WOOLDRIDGE (2022): "Simple alternatives to the common correlated effects model," ArXiv:2112.01486.

CHAN, M. K. AND S. S. KWOK (2022): "The PCDID approach: difference-in-differences when trends are potentially unparallel and stochastic," *Journal of Business & Economic Statistics*, 40, 1216–1233.

PESARAN, M. H. (2006): "Estimation and inference in large heterogeneous panels with a multifactor error structure," *Econometrica*, 74, 967–1012.

WESTERLUND, J., Y. PETROVA, AND M. NORKUTE (2019): "CCE in fixed-T panels," *Journal of Applied Econometrics*, 34, 746–761.