

ONLINE APPENDIX TO “DIRECT AND INDIRECT TREATMENT EFFECTS WITH TIME-VARYING COVARIATES”

Nicholas L. Brown
Florida State University

Kyle Butts
University of Arkansas

Joakim Westerlund*
Lund University
and
Deakin University

December 9, 2024

Abstract

This appendix provides (1) the proofs of Theorems 1 and 2 reported in the main paper, (2) a detailed account of the presence of covariates without common factor representation, and (3) the details of the bootstrap approach mentioned in the main paper.

1 Proofs

Proof of Theorem 1.

We first note that the proof in this appendix contains a slight generalization of the main results. Instead of assuming common slopes β , we assume unit-specific slopes β_i that satisfy the following random effects condition:

$$\beta_i = \beta + \eta_i \tag{A.1}$$

where η_i is mean zero and independent of all random components in the model. Such an assumption is common in the factor model literature.

*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.

We begin by considering the step-1 estimator of \mathbf{f}_t . In so doing, it is useful to let $N_g := |\mathcal{I}_g|$ for $g \in \{0, \dots, G\}$, just as in the main paper, and to denote by $\bar{\mathbf{a}}_t := N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{a}_{i,t}$ the cross-sectional average of any vector $\mathbf{a}_{i,t}$ for the group of untreated units ($g = 0$). In this notation, $\hat{\mathbf{f}}_t = \bar{\mathbf{z}}_t$. By combining the models for $y_{i,t}$ and $\mathbf{x}_{i,t}$ for untreated units, we arrive at the following static factor model for $\mathbf{z}_{i,t}$:

$$\mathbf{z}_{i,t} = \Lambda_i' \mathbf{f}_t + \mathbf{e}_{i,t}, \quad (\text{A.2})$$

where $\Lambda_i := [\alpha_i + \lambda_i \beta_i, \lambda_i]$ is $r \times (m+1)$ and $\mathbf{e}_{i,t} := [\varepsilon_{i,t} + \beta_i' \mathbf{v}_{i,t}, \mathbf{v}_{i,t}']'$ is $(m+1) \times 1$. This expression for $\mathbf{z}_{i,t}$ implies that $\hat{\mathbf{f}}_t$ can be written in the following way:

$$\hat{\mathbf{f}}_t = \bar{\mathbf{z}}_t = \bar{\Lambda}' \mathbf{f}_t + \bar{\mathbf{e}}_t. \quad (\text{A.3})$$

Here, $\bar{\Lambda}$ and $\bar{\mathbf{e}}_t$ are the cross-sectional averages of Λ_i and $\mathbf{e}_{i,t}$, respectively. For later use it is convenient to partition the first matrix as $\Lambda_i = [\alpha_i + \lambda_i \beta_i, \lambda_i] =: [\Lambda_{y,i}, \Lambda_{x,i}]$ with obvious definitions of $\Lambda_{y,i}$ and $\Lambda_{x,i}$. If $m+1 = r$, then the $r \times (m+1)$ matrix $\bar{\Lambda}$ is square and invertible, which means that (A.3) can be rewritten as

$$\bar{\Lambda}^{-1} \hat{\mathbf{f}}_t = \mathbf{f}_t + \bar{\Lambda}^{-1} \bar{\mathbf{e}}_t. \quad (\text{A.4})$$

Hence, because $\|\bar{\mathbf{e}}_t\| = O_p(N_0^{-1/2})$ under Assumption 6, we have

$$\bar{\Lambda}^{-1} \hat{\mathbf{f}}_t = \mathbf{f}_t + O_p(N_0^{-1/2}) \quad (\text{A.5})$$

and hence $\bar{\Lambda}^{-1} \hat{\mathbf{f}}_t$ is consistent for \mathbf{f}_t . In practice, we never observe $\bar{\Lambda}$. However, since $\alpha_i' \mathbf{f}_t = \alpha_i' \bar{\Lambda}^{-1} \hat{\mathbf{f}}_t + O_p(N_0^{-1/2})$, it is enough if we know $\hat{\mathbf{f}}_t$, because $\bar{\Lambda}^{-1}$ is subsumed in the estimation of the coefficient of $\hat{\mathbf{f}}_t$, which is \mathbf{a}_i in our notation.

The above analysis is not possible when $m+1 > r$ since $\bar{\Lambda}$ is no longer invertible. However, we still need something similar to (A.4), 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 $\bar{\Lambda}$ as $\bar{\Lambda} =: [\bar{\Lambda}_r, \bar{\Lambda}_{-r}]$, where $\bar{\Lambda}_{-r}$ is $r \times (m+1-r)$ and $\bar{\Lambda}_r$ is $r \times r$ and full rank. Note that this partition is without loss of generality under Assumption 7. 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 $m+1=r$:

$$\overline{\mathbf{H}} := \begin{bmatrix} \overline{\Lambda}_r^{-1} & -\overline{\Lambda}_r^{-1}\overline{\Lambda}_{-r} \\ \mathbf{0}_{(m+1-r) \times r} & \mathbf{I}_{m+1-r} \end{bmatrix} = [\overline{\mathbf{H}}_r, \overline{\mathbf{H}}_{-r}], \quad (\text{A.6})$$

where $\overline{\mathbf{H}}_r := [\overline{\Lambda}_r^{-1}, \mathbf{0}_{r \times (m+1-r)}]'$ is $(m+1) \times r$, while $\overline{\mathbf{H}}_{-r} := [-\overline{\Lambda}_{-r}'\overline{\Lambda}_r^{-1}, \mathbf{I}_{m+1-r}]'$ 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 $\mathbf{D}_{N_0} := \text{diag}(\mathbf{I}_r, \sqrt{N_0}\mathbf{I}_{m+1-r})$ with $\mathbf{D}_{N_0} := \mathbf{I}_{m+1}$ if $m+1=r$. By pre-multiplying $\widehat{\mathbf{f}}_t$ by $\mathbf{D}_{N_0}\overline{\mathbf{H}}'$, we obtain

$$\mathbf{D}_{N_0}\overline{\mathbf{H}}'\widehat{\mathbf{f}}_t = \widehat{\mathbf{f}}_t^0 = \mathbf{D}_{N_0}\overline{\mathbf{H}}'\overline{\Lambda}'\mathbf{f}_t + \mathbf{D}_{N_0}\overline{\mathbf{H}}'\overline{\mathbf{e}}_t = \mathbf{f}_t^0 + \overline{\mathbf{e}}_t^0, \quad (\text{A.7})$$

where $\mathbf{f}_t^0 := [\mathbf{f}_t', \mathbf{0}_{(m+1-r) \times 1}']'$ and $\overline{\mathbf{e}}_t^0 := [\overline{\mathbf{e}}_t'\overline{\mathbf{H}}_r, \sqrt{N_0}\overline{\mathbf{e}}_t'\overline{\mathbf{H}}_{-r}]' =: [\overline{\mathbf{e}}_{r,t}^{0'}, \overline{\mathbf{e}}_{-r,t}^{0'}]'$ are both $(m+1) \times 1$ with $\overline{\mathbf{e}}_{r,t}^{0'}$ and $\overline{\mathbf{e}}_{-r,t}^{0'}$ being $r \times 1$ and $(m+1-r) \times 1$, respectively. Hence, since $\|\overline{\mathbf{e}}_{r,t}^{0'}\| = O_p(N_0^{-1/2})$ and $\|\overline{\mathbf{e}}_{-r,t}^{0'}\| = O_p(1)$, when $m+1 > r$ we are no longer estimating \mathbf{f}_t but rather $\mathbf{f}_t^+ := [\mathbf{f}_t', \overline{\mathbf{e}}_{-r,t}^{0'}]'$;

$$\widehat{\mathbf{f}}_t^0 = \mathbf{f}_t^0 + \overline{\mathbf{e}}_t^0 = \begin{bmatrix} \mathbf{f}_t \\ \mathbf{0}_{(m+1-r) \times 1} \end{bmatrix} + \begin{bmatrix} \overline{\mathbf{e}}_{r,t}^{0'} \\ \overline{\mathbf{e}}_{-r,t}^{0'} \end{bmatrix} = \mathbf{f}_t^+ + O_p(N_0^{-1/2}). \quad (\text{A.8})$$

The fact that \mathbf{f}_t is included in \mathbf{f}_t^+ suggests that asymptotically TECCE should be able to account for the unknown factors even if $m+1 > r$. By ensuring the existence of $\overline{\mathbf{H}}$, Assumption 7 makes this possible. However, we also note that because of the presence of $\overline{\mathbf{e}}_{-r,t}^{0'}$, 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 $\widehat{\mathbf{f}}_t^0 := \overline{\Lambda}^{-1}\widehat{\mathbf{f}}_t$, $\mathbf{f}_t^0 := \mathbf{f}_t$ and $\overline{\mathbf{e}}_t^0 := \overline{\Lambda}^{-1}\overline{\mathbf{e}}_t$ if $m+1=r$, so that we are back in (A.4).

Let us now consider $\widehat{\Delta}_{i,t}$, which, unlike $\widehat{\mathbf{f}}_t$, is computed based on treated units in post-treatment periods ($i \in \mathcal{I}_0^c$ and $t > T_0$). We do our analysis for a given group g . 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 $\Delta_{i,t}$, $y_{i,t}(0)$ and $\tau_{i,t}$ leads to the following model

for $y_{i,t}$:

$$\begin{aligned}
y_{i,t} &= \Delta_{i,t} + y_{i,t}(0) \\
&= \Delta_{i,t} + \beta'_i \mathbf{x}_{i,t}(0) + \alpha'_i \mathbf{f}_t + \varepsilon_{i,t} \\
&= \Delta_{i,t} + \beta'_i (\mathbf{x}_{i,t} - \tau_{i,t}) + \alpha'_i \mathbf{f}_t + \varepsilon_{i,t} \\
&= (\Delta_{i,t} - \beta'_i \tau_{i,t}) + \beta'_i \mathbf{x}_{i,t} + \alpha'_i \mathbf{f}_t + \varepsilon_{i,t} \\
&= \eta_{i,t} + \beta'_i \mathbf{x}_{i,t} + \alpha'_i \mathbf{f}_t + \varepsilon_{i,t}.
\end{aligned} \tag{A.9}$$

It follows that

$$\begin{aligned}
\widehat{\Delta}_{i,t} &= y_{i,t} - \widehat{y}_{i,t}(0) \\
&= \eta_{i,t} + \beta'_i \mathbf{x}_{i,t} + \alpha'_i \mathbf{f}_t + \varepsilon_{i,t} - \widehat{\mathbf{a}}_i' \widehat{\mathbf{f}}_t \\
&= \eta_{i,t} + \beta'_i \mathbf{x}_{i,t} - (\widehat{\mathbf{a}}_i' \widehat{\mathbf{f}}_t - \alpha'_i \mathbf{f}_t) + \varepsilon_{i,t}.
\end{aligned} \tag{A.10}$$

Consider $\widehat{\mathbf{a}}_i' \widehat{\mathbf{f}}_t - \alpha'_i \mathbf{f}_t$. While the $(m+1) \times r$ matrix $\mathbf{D}_{N_0} \bar{\mathbf{H}}' \bar{\Lambda}'$ is not necessarily square under Assumption 7, it has full column rank. This means that we can compute its Moore–Penrose inverse, which is given by $(\mathbf{D}_{N_0} \bar{\mathbf{H}}' \bar{\Lambda}')^+ = (\mathbf{D}_{N_0} \bar{\mathbf{H}}' \bar{\Lambda}')' = [\mathbf{I}_r, \mathbf{0}_{r \times (m+1-r)}]$, such that $(\mathbf{D}_{N_0} \bar{\mathbf{H}}' \bar{\Lambda}')^+ \mathbf{D}_{N_0} \bar{\mathbf{H}}' \bar{\Lambda}' = \mathbf{I}_r$. Hence, $\mathbf{D}_{N_0} \bar{\mathbf{H}}' \bar{\Lambda}' \mathbf{f}_t = [\mathbf{f}_t', \mathbf{0}'_{(m+1-r) \times 1}]' = \mathbf{f}_t^0$ and we also have $\mathbf{D}_{N_0} \bar{\mathbf{H}}' \widehat{\mathbf{f}}_t = \widehat{\mathbf{f}}_t^0$. Making use of this, and letting $\widehat{\mathbf{a}}_i^0 := (\mathbf{D}_{N_0} \bar{\mathbf{H}}')^{-1} \widehat{\mathbf{a}}_i = (\bar{\mathbf{H}} \mathbf{D}_{N_0})^{-1} \widehat{\mathbf{a}}_i$ and $\alpha_i^0 := (\mathbf{D}_{N_0} \bar{\mathbf{H}}' \bar{\Lambda}')^+ \alpha_i = \mathbf{D}_{N_0} \bar{\mathbf{H}}' \bar{\Lambda}' \alpha_i = [\alpha_i', \mathbf{0}_{1 \times (m+1-r)}]',$ we obtain

$$\begin{aligned}
\widehat{\mathbf{a}}_i' \widehat{\mathbf{f}}_t - \alpha'_i \mathbf{f}_t &= \widehat{\mathbf{a}}_i' (\mathbf{D}_{N_0} \bar{\mathbf{H}}')^{-1} \mathbf{D}_{N_0} \bar{\mathbf{H}}' \widehat{\mathbf{f}}_t - \alpha'_i (\mathbf{D}_{N_0} \bar{\mathbf{H}}' \bar{\Lambda}')^+ \mathbf{D}_{N_0} \bar{\mathbf{H}}' \bar{\Lambda}' \mathbf{f}_t \\
&= \widehat{\mathbf{a}}_i^0 \widehat{\mathbf{f}}_t^0 - \alpha_i^0 \mathbf{f}_t^0 \\
&= \alpha_i^0 (\widehat{\mathbf{f}}_t^0 - \mathbf{f}_t^0) + (\widehat{\mathbf{a}}_i^0 - \alpha_i^0) \widehat{\mathbf{f}}_t^0,
\end{aligned} \tag{A.11}$$

from which it follows that

$$\widehat{\Delta}_{i,t} = \eta_{i,t} + \beta'_i \mathbf{x}_{i,t} - \alpha_i^0 (\widehat{\mathbf{f}}_t^0 - \mathbf{f}_t^0) - (\widehat{\mathbf{a}}_i^0 - \alpha_i^0) \widehat{\mathbf{f}}_t^0 + \varepsilon_{i,t}. \tag{A.12}$$

Amongst the terms appearing on the right-hand side of this last equation, the one involving $\widehat{\mathbf{a}}_i^0 - \alpha_i^0$ 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 $y_{i,t} = y_{i,t}(0) = \beta'_i \mathbf{x}_{i,t}(0) + \alpha'_i \mathbf{f}_t + \varepsilon_{i,t}$ or,

in terms of the stacked vector notation introduced in Section 3 of 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 T_0 -rowed. By using this and $\overline{\Lambda} \overline{\mathbf{H}}_r = \mathbf{I}_r$, we get

$$\mathbf{y}_i = \mathbf{x}_i\boldsymbol{\beta}_i + \widehat{\mathbf{f}}\overline{\mathbf{H}}_r\boldsymbol{\alpha}_i - (\widehat{\mathbf{f}} - \mathbf{f}\overline{\Lambda})\overline{\mathbf{H}}_r\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i = \mathbf{x}_i\boldsymbol{\beta}_i + \widehat{\mathbf{f}}\overline{\mathbf{H}}_r\boldsymbol{\alpha}_i - \bar{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i. \quad (\text{A.13})$$

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.13) into the expression given for $\widehat{\mathbf{a}}_i$ in step 2,

$$\begin{aligned} \widehat{\mathbf{a}}_i &= (\widehat{\mathbf{f}}\widehat{\mathbf{f}})^{-1}\widehat{\mathbf{f}}'\mathbf{y}_i \\ &= (\widehat{\mathbf{f}}\widehat{\mathbf{f}})^{-1}\widehat{\mathbf{f}}'(\mathbf{x}_i\boldsymbol{\beta}_i + \widehat{\mathbf{f}}\boldsymbol{\alpha}_i - \bar{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i) \\ &= \mathbf{a}_i + (\widehat{\mathbf{f}}\widehat{\mathbf{f}})^{-1}\widehat{\mathbf{f}}'(\mathbf{x}_i\boldsymbol{\beta}_i - \bar{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i), \end{aligned} \quad (\text{A.14})$$

implying

$$\begin{aligned} \widehat{\mathbf{a}}_i^0 &= (\overline{\mathbf{H}}\mathbf{D}_{N_0})^{-1}\widehat{\mathbf{a}}_i \\ &= (\overline{\mathbf{H}}\mathbf{D}_{N_0})^{-1}\mathbf{a}_i + (\overline{\mathbf{H}}\mathbf{D}_{N_0})^{-1}(\widehat{\mathbf{f}}\widehat{\mathbf{f}})^{-1}\widehat{\mathbf{f}}'(\mathbf{x}_i\boldsymbol{\beta}_i - \bar{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i) \\ &= (\overline{\mathbf{H}}\mathbf{D}_{N_0})^{-1}\mathbf{a}_i + (\mathbf{D}_{N_0}\overline{\mathbf{H}}'\widehat{\mathbf{f}}\widehat{\mathbf{f}}\overline{\mathbf{H}}\mathbf{D}_{N_0})^{-1}\mathbf{D}_{N_0}\overline{\mathbf{H}}'\widehat{\mathbf{f}}'(\mathbf{x}_i\boldsymbol{\beta}_i - \bar{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i) \\ &= (\overline{\mathbf{H}}\mathbf{D}_{N_0})^{-1}\mathbf{a}_i + (\widehat{\mathbf{f}}^0\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}^{0\prime}(\mathbf{x}_i\boldsymbol{\beta}_i - \bar{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i) \end{aligned} \quad (\text{A.15})$$

where $\widehat{\mathbf{f}}^0 = [\widehat{\mathbf{f}}_1^0, \dots, \widehat{\mathbf{f}}_{T_0}^0]':=\widehat{\mathbf{f}}\overline{\mathbf{H}}\mathbf{D}_{N_0}$ is $T_0 \times (m+1)$. 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

$$(\mathbf{D}_{N_0}\overline{\mathbf{H}})^{-1} = \begin{bmatrix} \overline{\Lambda}_r & \overline{\Lambda}_{-r} \\ \mathbf{0}_{(m+1-r) \times r} & N_0^{-1/2}\mathbf{I}_{m+1-r} \end{bmatrix}, \quad (\text{A.16})$$

so that

$$\begin{aligned} (\overline{\mathbf{H}}\mathbf{D}_{N_0})^{-1}\overline{\mathbf{H}}_r &= \begin{bmatrix} \overline{\Lambda}_r & \overline{\Lambda}_{-r} \\ \mathbf{0}_{(m+1-r) \times r} & N_0^{-1/2}\mathbf{I}_{m+1-r} \end{bmatrix} \begin{bmatrix} \overline{\Lambda}_r^{-1} \\ \mathbf{0}_{(m+1-r) \times r} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_r \\ \mathbf{0}_{(m+1-r) \times r} \end{bmatrix}. \end{aligned} \quad (\text{A.17})$$

This implies

$$(\overline{\mathbf{H}}\mathbf{D}_{N_0})^{-1}\mathbf{a}_i = \begin{bmatrix} \boldsymbol{\alpha}_i \\ \mathbf{0}_{(m+1-r) \times 1} \end{bmatrix} = \boldsymbol{\alpha}_i^0, \quad (\text{A.18})$$

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

$$\widehat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_i^0 = (\widehat{\mathbf{f}}^0/\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}^0'(\mathbf{x}_i\boldsymbol{\beta}_i - \bar{\mathbf{e}}_r^0\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i). \quad (\text{A.19})$$

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

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

Further use of $\widehat{\mathbf{f}} = \widehat{\mathbf{f}}^0 \mathbf{D}_{N_0}^{-1} \bar{\mathbf{H}}^{-1}$ gives

$$\mathbf{x}_i = \mathbf{f}\lambda_i + \mathbf{v}_i = \widehat{\mathbf{f}}\bar{\mathbf{H}}_r\lambda_i - (\widehat{\mathbf{f}} - \mathbf{f}\bar{\Lambda})\bar{\mathbf{H}}_r\lambda_i + \mathbf{v}_i = \widehat{\mathbf{f}}^0 \mathbf{D}_{N_0}^{-1} \bar{\mathbf{H}}^{-1} \bar{\mathbf{H}}_r \lambda_i - \bar{\mathbf{e}}_r^0 \lambda_i + \mathbf{v}_i, \quad (\text{A.21})$$

for pretreatment periods ($t \leq T_0$). If, on the other hand, $t > T_0$, then

$$\mathbf{x}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\lambda}'_i \mathbf{f}_t + \mathbf{v}_{i,t} = \boldsymbol{\tau}_{i,t} + \boldsymbol{\lambda}'_i \bar{\mathbf{H}}'_r \bar{\mathbf{H}}^{-1'} \mathbf{D}_{N_0}^{-1} \widehat{\mathbf{f}}_t^0 - \boldsymbol{\lambda}'_i \bar{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t}. \quad (\text{A.22})$$

These two last results imply

$$\begin{aligned} \mathbf{x}_{i,t} - \mathbf{x}_i' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0/\widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 \\ &= \boldsymbol{\tau}_{i,t} + \boldsymbol{\lambda}'_i \bar{\mathbf{H}}'_r \bar{\mathbf{H}}^{-1'} \mathbf{D}_{N_0}^{-1} \widehat{\mathbf{f}}_t^0 - \boldsymbol{\lambda}'_i \bar{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t} - (\widehat{\mathbf{f}}^0 \mathbf{D}_{N_0}^{-1} \bar{\mathbf{H}}^{-1} \bar{\mathbf{H}}_r \lambda_i - \bar{\mathbf{e}}_r^0 \lambda_i + \mathbf{v}_i)' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0/\widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 \\ &= \boldsymbol{\tau}_{i,t} - \boldsymbol{\lambda}'_i \bar{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t} - (-\bar{\mathbf{e}}_r^0 \lambda_i + \mathbf{v}_i)' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0/\widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0, \end{aligned} \quad (\text{A.23})$$

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

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

where we recall that $\boldsymbol{\Lambda}_{y,i} := \boldsymbol{\lambda}_i \boldsymbol{\beta}_i + \boldsymbol{\alpha}_i$ and $\Delta_{i,t} = \eta_{i,t} + \boldsymbol{\beta}'_i \boldsymbol{\tau}_{i,t}$ as in the main paper.

The above expression for $\widehat{\Delta}_{i,t}$ is the cleanest possible without exploiting the fact that N is large. Hence, in what remains we are going to let $N \rightarrow \infty$. We begin by considering $\widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0/\widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0$. Recall that $\mathbf{f}^+ = [\mathbf{f}_1^+, \dots, \mathbf{f}_{T_0}^+]' := [\mathbf{f}, \bar{\mathbf{e}}_{-r}^0]$, a $T_0 \times (m+1)$ matrix. We have already shown that $\widehat{\mathbf{f}}^0 = \mathbf{f}^+ + O_p(N_0^{-1/2})$. By using this and the results provided in the proof of

Lemma A.1 in Westerlund et al. (2019), we obtain $\|\widehat{\mathbf{f}}^0\widehat{\mathbf{f}}^0 - \mathbf{f}^{+'}\mathbf{f}^+\| = O_p(N_0^{-1/2})$ and, more importantly,

$$\|(\widehat{\mathbf{f}}^0\widehat{\mathbf{f}}^0)^{-1} - (\mathbf{f}^{+'}\mathbf{f}^+)^{-1}\| = O_p(N_0^{-1/2}), \quad (\text{A.25})$$

where

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

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

The expression for $(\mathbf{f}^{+'}\mathbf{f}^+)^{-1}$ is again obtained by using the rules for the inverse of a partitioned matrix. Combining $\|(\widehat{\mathbf{f}}^0\widehat{\mathbf{f}}^0)^{-1} - (\mathbf{f}^{+'}\mathbf{f}^+)^{-1}\| = O_p(N_0^{-1/2})$ with $\widehat{\mathbf{f}}^0 = \mathbf{f}^+ + O_p(N_0^{-1/2})$, we obtain

$$\begin{aligned} \widehat{\mathbf{f}}_t^{0'}(\widehat{\mathbf{f}}^0\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}^{0'} &= \widehat{\mathbf{f}}_t^{0'}[(\widehat{\mathbf{f}}^0\widehat{\mathbf{f}}^0)^{-1} - (\mathbf{f}^{+'}\mathbf{f}^+)^{-1}]\widehat{\mathbf{f}}^{0'} + \widehat{\mathbf{f}}_t^{0'}(\mathbf{f}^{+'}\mathbf{f}^+)^{-1}\widehat{\mathbf{f}}^{0'} \\ &= \widehat{\mathbf{f}}_t^{0'}(\mathbf{f}^{+'}\mathbf{f}^+)^{-1}\widehat{\mathbf{f}}^{0'} + O_p(N_0^{-1/2}) \\ &= \mathbf{f}_t^{+'}(\mathbf{f}^{+'}\mathbf{f}^+)^{-1}\mathbf{f}^{+'} + O_p(N_0^{-1/2}). \end{aligned} \quad (\text{A.28})$$

where, defining \mathbf{M}_f analogously to $\mathbf{M}_{\widehat{f}}$,

$$\begin{aligned} &\mathbf{f}_t^{+'}(\mathbf{f}^{+'}\mathbf{f}^+)^{-1}\mathbf{f}^{+'} \\ &= [\mathbf{f}_t', \bar{\mathbf{e}}_{-r,t}^{0'}] \begin{bmatrix} (\mathbf{f}'\mathbf{f})^{-1} + (\mathbf{f}'\mathbf{f})^{-1}\mathbf{f}'\bar{\mathbf{e}}_{-r}^0(\bar{\mathbf{e}}_{-r}^{0'}\mathbf{M}_f\bar{\mathbf{e}}_{-r}^0)^{-1}\bar{\mathbf{e}}_{-r}^{0'}\mathbf{f}(\mathbf{f}'\mathbf{f})^{-1} \\ -(\bar{\mathbf{e}}_{-r}^{0'}\mathbf{M}_f\bar{\mathbf{e}}_{-r}^0)^{-1}\bar{\mathbf{e}}_{-r}^{0'}\mathbf{f}(\mathbf{f}'\mathbf{f})^{-1} \\ -(\mathbf{f}'\mathbf{f})^{-1}\mathbf{f}'\bar{\mathbf{e}}_{-r}^0(\bar{\mathbf{e}}_{-r}^{0'}\mathbf{M}_f\bar{\mathbf{e}}_{-r}^0)^{-1} \\ (\bar{\mathbf{e}}_{-r}^{0'}\mathbf{M}_f\bar{\mathbf{e}}_{-r}^0)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{f}' \\ \bar{\mathbf{e}}_{-r}^{0'} \end{bmatrix} \\ &= \mathbf{f}_t'(\mathbf{f}'\mathbf{f})^{-1}\mathbf{f}'[\mathbf{I}_{T_0} - \bar{\mathbf{e}}_{-r}^0(\bar{\mathbf{e}}_{-r}^{0'}\mathbf{M}_f\bar{\mathbf{e}}_{-r}^0)^{-1}\bar{\mathbf{e}}_{-r}^{0'}\mathbf{M}_f] + \bar{\mathbf{e}}_{-r,t}^{0'}(\bar{\mathbf{e}}_{-r}^{0'}\mathbf{M}_f\bar{\mathbf{e}}_{-r}^0)^{-1}\bar{\mathbf{e}}_{-r}^{0'}\mathbf{M}_f. \end{aligned} \quad (\text{A.29})$$

Note that this last equation reduces to $\mathbf{f}_t^{+'}(\mathbf{f}^{+'}\mathbf{f}^+)^{-1}\mathbf{f}^{+'} = \mathbf{f}_t'(\mathbf{f}'\mathbf{f})^{-1}\mathbf{f}'$ if $m+1=r$. Hence, as mentioned earlier, the asymptotic distribution theory depends on whether $m+1=r$ or $m+1>r$.

Use of $\|\widehat{\mathbf{f}}_t^{0'}(\widehat{\mathbf{f}}^0\widehat{\mathbf{f}}^0)^{-1}\widehat{\mathbf{f}}^{0'} - \mathbf{f}_t^{+'}(\mathbf{f}^{+'}\mathbf{f}^+)^{-1}\mathbf{f}^{+'}\| = O_p(N_0^{-1/2})$ leads to the following for $\widehat{\Delta}_{i,t}$:

$$\begin{aligned} \widehat{\Delta}_{i,t} &= \Delta_{i,t} - \boldsymbol{\Lambda}'_{y,i}\bar{\mathbf{e}}_{r,t}^0 + \boldsymbol{\beta}'_i\mathbf{v}_{i,t} + \varepsilon_{i,t} - (-\bar{\mathbf{e}}^0\boldsymbol{\Lambda}_{y,i} + \mathbf{v}_i\boldsymbol{\beta}_i + \varepsilon_i)' \mathbf{f}^+(\mathbf{f}^{+'}\mathbf{f}^+)^{-1}\mathbf{f}_t^+ + O_p(N_0^{-1/2}) \\ &= \Delta_{i,t} - \boldsymbol{\Lambda}'_{y,i}\bar{\mathbf{e}}_{r,t}^{0d} + \boldsymbol{\beta}'_i\mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d + O_p(N_0^{-1/2}), \end{aligned} \quad (\text{A.30})$$

where

$$\mathbf{a}_{i,t}^d = \mathbf{a}_{i,t} - \mathbf{a}_i' \mathbf{f}^+ (\mathbf{f}^{+'} \mathbf{f}^+)^{-1} \mathbf{f}_t^+ = \mathbf{a}_{i,t} - \sum_{s=1}^{T_0} \mathbf{a}_{i,s} \mathbf{f}_s^{+'} (\mathbf{f}^{+'} \mathbf{f}^+)^{-1} \mathbf{f}_t^+ \quad (\text{A.31})$$

for any vector $\mathbf{a}_{i,t}$ with T_0 -rowed stack $\mathbf{a}_i = [\mathbf{a}_{i,1}, \dots, \mathbf{a}_{i,T_0}]'$. In words, $\mathbf{a}_{i,t}^d$ is the limiting “defactored” version of $\mathbf{a}_{i,t}$. It is important to note that the defactoring leaves properties of $\bar{\mathbf{e}}_{r,t}^0, \mathbf{v}_{i,t}$ and $\varepsilon_{i,t}$ basically unaffected. This is so because $\{(\bar{\mathbf{e}}_{r,t}^0, \mathbf{v}_{i,t}, \varepsilon_{i,t})\}_{i \in \mathcal{I}_0^c}$ and the sequence $\{\mathbf{e}_{j,t}^d\}_{j \in \mathcal{I}_0}$ sitting in \mathbf{f}^+ are from non-overlapping subsamples. In fact, the only difference is the variance (conditional on \mathbf{f}^+). As an example, consider $\varepsilon_{i,t}^d$. Denoting by \mathcal{C} the sigma-field generated by \mathbf{f}^+ , we have

$$\begin{aligned} \mathbb{E}(\varepsilon_{i,t}^{d2} | \mathcal{C}) &= \mathbb{E}[(\varepsilon_{i,t} - \mathbf{e}_i' \mathbf{f}^+ (\mathbf{f}^{+'} \mathbf{f}^+)^{-1} \mathbf{f}_t^+)^2 | \mathcal{C}] \\ &= \mathbb{E}(\varepsilon_{i,t}^2 | \mathcal{C}) - 2\mathbb{E}(\varepsilon_{i,t} \mathbf{e}_i' | \mathcal{C}) \mathbf{f}^+ (\mathbf{f}^{+'} \mathbf{f}^+)^{-1} \mathbf{f}_t^+ \\ &\quad + \mathbf{f}_t^{+'} (\mathbf{f}^{+'} \mathbf{f}^+)^{-1} \mathbf{f}^{+'} \mathbb{E}(\varepsilon_i \mathbf{e}_i' | \mathcal{C}) \mathbf{f}^+ (\mathbf{f}^{+'} \mathbf{f}^+)^{-1} \mathbf{f}_t^+ \\ &= \sigma_{\varepsilon_{i,t}}^2 - 2\mathbf{1}_t' \boldsymbol{\Sigma}_{\varepsilon,i} \mathbf{f}^+ (\mathbf{f}^{+'} \mathbf{f}^+)^{-1} \mathbf{f}_t^+ + \mathbf{f}_t^{+'} (\mathbf{f}^{+'} \mathbf{f}^+)^{-1} \mathbf{f}^{+'} \boldsymbol{\Sigma}_{\varepsilon,i} \mathbf{f}^+ (\mathbf{f}^{+'} \mathbf{f}^+)^{-1} \mathbf{f}_t^+, \end{aligned} \quad (\text{A.32})$$

where $\sigma_{\varepsilon_{i,t}}^2 = \mathbb{E}(\varepsilon_{i,t}^2)$, $\boldsymbol{\Sigma}_{\varepsilon,i} = \mathbb{E}(\varepsilon_i \varepsilon_i')$ and $\mathbf{1}_t$ is a $T_0 \times 1$ vector of zeros except for a one in row t .

We now make use of the above expression for $\widehat{\Delta}_{i,t}$ to evaluate $\widehat{\text{ATT}}_{g,t}$. In so doing, it is important to note that the order of the reminder incurred when replacing $\widehat{\mathbf{f}}_t^{0\prime} (\widehat{\mathbf{f}}^0 \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}^{0\prime}$ with $\mathbf{f}_t^{+'} (\mathbf{f}^{+'} \mathbf{f}^+)^{-1} \mathbf{f}^{+'}$ is the same even after averaging over group g and multiplying by $\sqrt{N_g}$. In order to appreciate this, we make use of the fact that $\|\sqrt{N_g} \bar{\mathbf{e}}_r^0\| = O_p(\sqrt{N_g/N_0}) = O_p(1)$ under $N_g/N_0 \rightarrow \tau < \infty$, and since \mathbf{v}_i and $\boldsymbol{\beta}_i$ are independent with \mathbf{v}_i mean zero and independent also across i , we also have $\|N_g^{-1/2} \sum_{i \in \mathcal{I}_g} \mathbf{v}_i \boldsymbol{\beta}_i\| = O_p(1)$. It follows that

$$\begin{aligned} &\left\| \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (-\bar{\mathbf{e}}_r^0 \boldsymbol{\Lambda}_{y,i} + \mathbf{v}_i \boldsymbol{\beta}_i + \varepsilon_i) \right\| \\ &\leq \|\sqrt{N_g} \bar{\mathbf{e}}_r^0\| \left\| \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \boldsymbol{\Lambda}_{y,i} \right\| + \left\| \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} \mathbf{v}_i \boldsymbol{\beta}_i \right\| + \left\| \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} \varepsilon_i \right\| \\ &= O_p(1). \end{aligned} \quad (\text{A.33})$$

We can therefore show that

$$\begin{aligned}
& \left\| \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (-\bar{\mathbf{e}}_r^0 \boldsymbol{\Lambda}_{y,i} + \mathbf{v}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i)' [\mathbf{f}^+ (\mathbf{f}^{+\prime} \mathbf{f}^+)^{-1} \mathbf{f}_t^+ - \hat{\mathbf{f}}^0 (\hat{\mathbf{f}}^{0\prime} \hat{\mathbf{f}}^0)^{-1} \hat{\mathbf{f}}_t^0] \right\| \\
& \leq \left\| \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} [-\bar{\mathbf{e}}_r^0 \boldsymbol{\Lambda}_{y,i} + \mathbf{v}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i] \right\| \|\mathbf{f}^+ (\mathbf{f}^{+\prime} \mathbf{f}^+)^{-1} \mathbf{f}_t^+ - \hat{\mathbf{f}}^0 (\hat{\mathbf{f}}^{0\prime} \hat{\mathbf{f}}^0)^{-1} \hat{\mathbf{f}}_t^0\| \\
& = O_p(N_0^{-1/2}),
\end{aligned} \tag{A.34}$$

which means that the reminder incurred when replacing $\hat{\mathbf{f}}_t^{0\prime} (\hat{\mathbf{f}}^{0\prime} \hat{\mathbf{f}}^0)^{-1} \hat{\mathbf{f}}_t^0$ with $\mathbf{f}_t^{+\prime} (\mathbf{f}^{+\prime} \mathbf{f}^+)^{-1} \mathbf{f}^+$ is $O_p(N_0^{-1/2})$ after averaging over group g and multiplying by $\sqrt{N_g}$.

For $\Delta_{i,t}$, we make use of the fact that $\Delta_{i,t} = \text{ATT}_{g,t} + v_{i,t}$ for $i \in \mathcal{I}_g$ by Assumption 1, giving

$$\begin{aligned}
\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) &= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (\widehat{\Delta}_{i,t} - \text{ATT}_{g,t}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (\widehat{\Delta}_{i,t} - \Delta_{i,t} + v_{i,t}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} - \boldsymbol{\Lambda}'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d} + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d + \boldsymbol{\varepsilon}_{i,t}^d) + O_p(N_0^{-1/2}).
\end{aligned} \tag{A.35}$$

Hence, if we in addition define $\boldsymbol{\Lambda}_y := \text{plim}_{N_g \rightarrow \infty} N_g^{-1} \sum_{i \in \mathcal{I}_g} \boldsymbol{\Lambda}_{y,i}$, the above expression for $\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t})$ becomes

$$\begin{aligned}
& \sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d + \boldsymbol{\varepsilon}_{i,t}^d) - \sqrt{\frac{N_g}{N_0}} \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \boldsymbol{\Lambda}'_{y,i} \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} + O_p(N_0^{-1/2}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d + \boldsymbol{\varepsilon}_{i,t}^d) - \sqrt{\tau} \boldsymbol{\Lambda}'_y \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} + o_p(1).
\end{aligned} \tag{A.36}$$

All the terms on the right-hand side of the above equation are mean zero and independent across i (conditionally on \mathcal{C}). They are therefore asymptotically normal by a central limit law for independent variables. Let us therefore define $\text{var}(\widehat{\text{ATT}}_{g,t}) = \text{var}[\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) | \mathcal{C}]$. The asymptotic distribution of $\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t})$ as $N_g, N_0 \rightarrow \infty$ with $N_g/N_0 \rightarrow \tau < \infty$ can now be stated in the following way:

$$\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) \xrightarrow{d} MN(0, \text{var}(\widehat{\text{ATT}}_{g,t})), \tag{A.37}$$

where $MN(\cdot, \cdot)$ signifies a mixed normal distribution that is normal conditionally on \mathcal{C} . This means that the conditional distribution of $\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) / \sqrt{\text{var}(\widehat{\text{ATT}}_{g,t})}$ is also the unconditional distribution. Hence,

$$\frac{\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t})}{\text{var}(\widehat{\text{ATT}}_{g,t})} \rightarrow_d N(0, 1), \quad (\text{A.38})$$

which is what we wanted to show.

In what remains of this proof we evaluate (the limit of) $\text{var}(\widehat{\text{ATT}}_{g,t})$. From before,

$$\begin{aligned} & \lim_{N_g, N_0 \rightarrow \infty} \text{var}(\widehat{\text{ATT}}_{g,t}) \\ &= \lim_{N_g, N_0 \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) - \sqrt{\tau} \boldsymbol{\Lambda}'_y \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} \right)^2 | \mathcal{C} \right], \end{aligned} \quad (\text{A.39})$$

where

$$\begin{aligned} & \mathbb{E} \left[\left(\frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) - \sqrt{\tau} \boldsymbol{\Lambda}'_y \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} \right)^2 | \mathcal{C} \right] \\ &= \mathbb{E} \left[\left(\frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) \right)^2 | \mathcal{C} \right] + \mathbb{E}[(\sqrt{\tau} \boldsymbol{\Lambda}'_y \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d})^2 | \mathcal{C}] \\ &\quad - 2\mathbb{E} \left(\frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) \sqrt{\tau} \boldsymbol{\Lambda}'_y \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} | \mathcal{C} \right) \\ &= \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \sum_{j \in \mathcal{I}_g} \mathbb{E}[(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d)(v_{j,t} + \beta'_j \mathbf{v}_{j,t}^d + \varepsilon_{j,t}^d) | \mathcal{C}] + \tau \boldsymbol{\Lambda}'_y N_0 \mathbb{E}(\bar{\mathbf{e}}_{r,t}^{0d} \bar{\mathbf{e}}_{r,t}^{0d\prime} | \mathcal{C}) \boldsymbol{\Lambda}_y \\ &\quad - 2\sqrt{\tau} \boldsymbol{\Lambda}'_y \sqrt{\frac{N_0}{N_g}} \sum_{i \in \mathcal{I}_g} \mathbb{E}[(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) \bar{\mathbf{e}}_{r,t}^{0d} | \mathcal{C}]. \end{aligned} \quad (\text{A.40})$$

We now consider each of the three terms on the right, starting with the first. Conditionally on \mathcal{C} , $v_{i,t}$, β_i , $\mathbf{v}_{i,t}^d$ and $\varepsilon_{i,t}^d$ are independent not only across i but also of one another. One implication of this is that

$$\begin{aligned} \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \mathbb{E}(\beta'_i \mathbf{v}_{i,t}^d \mathbf{v}_{i,t}^{d\prime} \beta_i | \mathcal{C}) &= \text{tr} \left(\frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \mathbb{E}(\beta_i \beta_i') \mathbb{E}(\mathbf{v}_{i,t}^d \mathbf{v}_{i,t}^{d\prime} | \mathcal{C}) \right) \\ &= \text{tr}(\boldsymbol{\Sigma}_\beta \boldsymbol{\Sigma}_{\mathbf{v}^d, t}) + o(1), \end{aligned} \quad (\text{A.41})$$

where $\Sigma_{\mathbf{v}^d,t} := \lim_{N_g \rightarrow \infty} N_g^{-1} \sum_{i \in \mathcal{I}_g} \mathbb{E}(\mathbf{v}_{i,t}^d \mathbf{v}_{i,t}^{d'} | \mathcal{C})$ and $\Sigma_{\beta} := \mathbb{E}(\beta_i \beta_i')$. Note that $\mathbb{E}(\mathbf{v}_{i,t}^d \mathbf{v}_{i,t}^{d'} | \mathcal{C})$ has the same basic structure as $\mathbb{E}(\varepsilon_{i,t}^{d2} | \mathcal{C})$ earlier. Another implication of the above mentioned conditional independence is that

$$\begin{aligned} & \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \sum_{j \in \mathcal{I}_g} \mathbb{E}[(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d)(v_{j,t} + \beta'_j \mathbf{v}_{j,t}^d + \varepsilon_{j,t}^d) | \mathcal{C}] \\ &= \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \mathbb{E}[(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d)^2 | \mathcal{C}] \\ &= \sigma_{v,t}^2 + \text{tr}(\Sigma_{\beta} \Sigma_{\mathbf{v}^d,t}) + \sigma_{\varepsilon^d,t}^2 + o(1), \end{aligned} \quad (\text{A.42})$$

where $\sigma_{v,t}^2 := \lim_{N_g \rightarrow \infty} N_g^{-1} \sum_{i \in \mathcal{I}_g} \mathbb{E}(v_{i,t}^2)$ and $\sigma_{\varepsilon^d,t}^2 := \lim_{N_g \rightarrow \infty} N_g^{-1} \sum_{i \in \mathcal{I}_g} \mathbb{E}(\varepsilon_{i,t}^{d2} | \mathcal{C})$. Similarly, because $\mathbf{e}_{i,t}^d$ is conditionally independent across i , defining $\Sigma_{\mathbf{e}^d,t}$ analogously to $\Sigma_{\mathbf{v}^d,t}$, we have

$$\begin{aligned} \mathbb{E}(\bar{\mathbf{e}}_{r,t}^{0d} \bar{\mathbf{e}}_{r,t}^{0d'} | \mathcal{C}) &= \bar{\mathbf{H}}_r' \mathbb{E}(\mathbf{e}_t^d \mathbf{e}_t^{d'} | \mathcal{C}) \bar{\mathbf{H}}_r \\ &= \bar{\mathbf{H}}_r' \frac{1}{N_0^2} \sum_{i \in \mathcal{I}_0} \sum_{j \in \mathcal{I}_0} \mathbb{E}(\mathbf{e}_{i,t}^d \mathbf{e}_{j,t}^{d'} | \mathcal{C}) \bar{\mathbf{H}}_r \\ &= \bar{\mathbf{H}}_r' \frac{1}{N_0^2} \sum_{i \in \mathcal{I}_0} \mathbb{E}(\mathbf{e}_{i,t}^d \mathbf{e}_{i,t}^{d'} | \mathcal{C}) \bar{\mathbf{H}}_r \\ &= N_0^{-1} \bar{\mathbf{H}}_r' \Sigma_{\mathbf{e}^d,t} \bar{\mathbf{H}}_r + o(1). \end{aligned} \quad (\text{A.43})$$

Hence, letting $\mathbf{H}_r = \lim_{N_0 \rightarrow \infty} \bar{\mathbf{H}}_r$, the second term on the right-hand side of (A.40) reduces to

$$\tau \Lambda_y' N_0 \mathbb{E}(\bar{\mathbf{e}}_{r,t}^{0d} \bar{\mathbf{e}}_{r,t}^{0d'} | \mathcal{C}) \Lambda_y = \tau \Lambda_y' \bar{\mathbf{H}}_r' \Sigma_{\mathbf{e}^d,t} \bar{\mathbf{H}}_r \Lambda_y + o(1) = \tau \Lambda_y' \mathbf{H}_r' \Sigma_{\mathbf{e}^d,t} \mathbf{H}_r \Lambda_y + o(1). \quad (\text{A.44})$$

The third and last term is zero. This is because $\{(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d)\}_{i \in \mathcal{I}_g}$ and $\{\mathbf{e}_{j,t}^d\}_{j \in \mathcal{I}_0}$ are mean zero and from non-overlapping subsamples, giving

$$\begin{aligned} & \sum_{i \in \mathcal{I}_g} \mathbb{E}[(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) \bar{\mathbf{e}}_{r,t}^{0d} | \mathcal{C}] \\ &= \bar{\mathbf{H}}_r' \frac{1}{N_0} \sum_{i \in \mathcal{I}_g} \sum_{j \in \mathcal{I}_0} \mathbb{E}[(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) \mathbf{e}_{j,t}^d | \mathcal{C}] = 0, \end{aligned} \quad (\text{A.45})$$

which in turn implies

$$\begin{aligned} & \lim_{N_g, N_0 \rightarrow \infty} \text{var}(\widehat{\text{ATT}}_{g,t}) \\ &= \lim_{N_g, N_0 \rightarrow \infty} \mathbb{E} \left[\left(\frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) - \sqrt{\tau} \Lambda_y' \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} \right)^2 | \mathcal{C} \right] \\ &= \sigma_{v,t}^2 + \text{tr}(\Sigma_{\beta} \Sigma_{\mathbf{v}^d,t}) + \sigma_{\varepsilon^d,t}^2 + \tau \Lambda_y' \mathbf{H}_r' \Sigma_{\mathbf{e}^d,t} \mathbf{H}_r \Lambda_y. \end{aligned} \quad (\text{A.46})$$

This completes the evaluation of $\text{var}(\widehat{\text{ATT}}_{g,t})$. The proof is therefore complete. \blacksquare

Proof of Theorem 2.

From the proof of Theorem 1, again with $\Lambda_{y,i} := \lambda_i \beta_i + \alpha_i$,

$$\widehat{\Delta}_{i,t} = \text{ATT}_{g,t} + v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d - \Lambda'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d} + O_p(N_0^{-1/2}), \quad (\text{A.47})$$

$$\frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \widehat{\Delta}_{i,t} = \text{ATT}_{g,t} + \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} [v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d - \Lambda'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d}] + O_p(N_0^{-1/2}). \quad (\text{A.48})$$

Hence, letting $z_{i,t} := v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d - \Lambda'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d}$, we obtain

$$\begin{aligned} \widehat{\Delta}_{i,t} - \frac{1}{N_g} \sum_{j \in \mathcal{I}_g} \widehat{\Delta}_{j,t} &= z_{i,t} - \frac{1}{N_g} \sum_{j \in \mathcal{I}_g} z_{j,t} + O_p(N_0^{-1/2}) \\ &= z_{i,t} + O_p(\min\{N_0, N_g\}^{-1/2}), \end{aligned} \quad (\text{A.49})$$

where the last equality holds because

$$\frac{1}{N_g} \sum_{j \in \mathcal{I}_g} z_{j,t} = \frac{1}{N_g} \sum_{j \in \mathcal{I}_g} (v_{j,t} + \beta'_j \mathbf{v}_{j,t}^d + \varepsilon_{j,t}^d - \Lambda'_{y,j} \bar{\mathbf{e}}_{r,t}^{0d}) = O_p(\min\{N_0, N_g\}^{-1/2}). \quad (\text{A.50})$$

It follows that

$$\begin{aligned} \widehat{\text{var}}(\widehat{\text{ATT}}_{g,t}) &= \frac{1}{N_g - 1} \sum_{i \in \mathcal{I}_g} \left(\widehat{\Delta}_{i,t} - \frac{1}{N_g} \sum_{j \in \mathcal{I}_g} \widehat{\Delta}_{j,t} \right)^2 \\ &= \frac{1}{N_g - 1} \sum_{i \in \mathcal{I}_g} z_{i,t}^2 + O_p(\min\{N_0, N_g\}^{-1/2}). \end{aligned} \quad (\text{A.51})$$

As for the remaining term, since again $v_{i,t}$, $\beta'_i \mathbf{v}_{i,t}^d$ and $\varepsilon_{i,t}^d$ are mean zero, and conditionally independent of each other as well as across i , their cross-products in $(N_g - 1)^{-1} \sum_{i \in \mathcal{I}_g} z_{i,t}^2$ are $O_p(N_g^{-1/2})$. All cross-products involving $\Lambda'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d}$ are $O_p((N_g N_0)^{-1/2})$, as is clear from the following example:

$$\left\| \frac{1}{N_g - 1} \sum_{i \in \mathcal{I}_g} v_{i,t} \Lambda'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d} \right\| \leq \left\| \frac{1}{N_g - 1} \sum_{i \in \mathcal{I}_g} v_{i,t} \Lambda_{y,i} \right\| \|\bar{\mathbf{e}}_{r,t}^{0d}\| = O_p((N_g N_0)^{-1/2}). \quad (\text{A.52})$$

Making use of this and

$$\frac{1}{N_g - 1} \sum_{i \in \mathcal{I}_g} \Lambda'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d} \bar{\mathbf{e}}_{r,t}^{0d'} \Lambda_{y,i} \leq \frac{1}{N_g - 1} \sum_{i \in \mathcal{I}_g} \|\Lambda_{y,i}\|^2 \|\bar{\mathbf{e}}_{r,t}^{0d}\|^2 = O_p(N_0^{-1}), \quad (\text{A.53})$$

we obtain

$$\begin{aligned}
\widehat{\text{var}}(\widehat{\text{ATT}}_{g,t}) &= \frac{1}{N_g - 1} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d - \boldsymbol{\Lambda}'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d})^2 + o_p(1) \\
&= \frac{1}{N_g - 1} \sum_{i \in \mathcal{I}_g} (v_{i,t}^2 + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d \mathbf{v}_{i,t}^{d'} \boldsymbol{\beta}_i + \varepsilon_{i,t}^{d2} + \boldsymbol{\Lambda}'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d} \bar{\mathbf{e}}_{r,t}^{0d'} \boldsymbol{\Lambda}_{y,i}) + o_p(1) \\
&= \frac{1}{N_g - 1} \sum_{i \in \mathcal{I}_g} (v_{i,t}^2 + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d \mathbf{v}_{i,t}^{d'} \boldsymbol{\beta}_i + \varepsilon_{i,t}^{d2}) + o_p(1) \\
&= \sigma_{v,t}^2 + \text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}_{\mathbf{v}^d,t}) + \sigma_{\varepsilon^d,t}^2 + o_p(1).
\end{aligned} \tag{A.54}$$

From Proof of Theorem 1, we have that

$$\text{var}(\widehat{\text{ATT}}_{g,t}) \rightarrow \sigma_{v,t}^2 + \text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}_{\mathbf{v}^d,t}) + \sigma_{\varepsilon^d,t}^2 + \tau \boldsymbol{\Lambda}'_y \mathbf{H}'_r \boldsymbol{\Sigma}_{\mathbf{e}^d,t} \mathbf{H}_r \boldsymbol{\Lambda}_y \tag{A.55}$$

as $N_g, N_0 \rightarrow \infty$ with $N_g/N_0 \rightarrow \tau < \infty$. It follows that under the additional assumption that $\tau = 0$, we have

$$\widehat{\text{var}}(\widehat{\text{ATT}}_{g,t}) \rightarrow_p \text{var}(\widehat{\text{ATT}}_{g,t}), \tag{A.56}$$

as required for the proof. ■

Proof of Theorem 3.

The proof of this theorem follows from simple manipulations of that of Theorem 1. We therefore only consider the main differences.

Let us start with the estimator of the DATT, $\widehat{\text{DATT}}_{g,t}$. We begin by writing $\widehat{\eta}_{i,t}$ in terms of $\eta_{i,t}$ and the primitives of the model. Unlike in Proof of Theorem 1, now $\widehat{y}_{i,t}(0) = \widehat{\boldsymbol{\beta}}' \mathbf{x}_{i,t} + \widehat{\mathbf{a}}'_i \widehat{\mathbf{f}}_t$. By using this and the same steps used to evaluate $\widehat{\Delta}_{i,t}$ in Proof of Theorem 1, we obtain

$$\begin{aligned}
\widehat{\eta}_{i,t} &= y_{i,t} - \widehat{y}_{i,t}(0) \\
&= \eta_{i,t} + \boldsymbol{\beta}'_i \mathbf{x}_{i,t} + \boldsymbol{\alpha}'_i \mathbf{f}_t + \varepsilon_{i,t} - (\widehat{\boldsymbol{\beta}}' \mathbf{x}_{i,t} + \widehat{\mathbf{a}}'_i \widehat{\mathbf{f}}_t) \\
&= \eta_{i,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' \mathbf{x}_{i,t} - (\widehat{\mathbf{a}}'_i \widehat{\mathbf{f}}_t - \boldsymbol{\alpha}'_i \mathbf{f}_t) + \varepsilon_{i,t} \\
&= \eta_{i,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' \mathbf{x}_{i,t} - \boldsymbol{\alpha}'_i (\widehat{\mathbf{f}}_t^0 - \mathbf{f}_t^0) - (\widehat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_i^0)' \widehat{\mathbf{f}}_t^0 + \varepsilon_{i,t}.
\end{aligned} \tag{A.57}$$

Here

$$\begin{aligned}
\widehat{\mathbf{a}}_i &= (\widehat{\mathbf{f}}' \widehat{\mathbf{f}})^{-1} \widehat{\mathbf{f}}' (\mathbf{y}_i - \mathbf{x}_i \widehat{\boldsymbol{\beta}}) \\
&= (\widehat{\mathbf{f}}' \widehat{\mathbf{f}})^{-1} \widehat{\mathbf{f}}' [-\mathbf{x}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) + \widehat{\mathbf{f}} \mathbf{a}_i - \bar{\mathbf{e}}_r^0 \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i] \\
&= \mathbf{a}_i + (\widehat{\mathbf{f}}' \widehat{\mathbf{f}})^{-1} \widehat{\mathbf{f}}' [-\mathbf{x}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) - \bar{\mathbf{e}}_r^0 \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i],
\end{aligned} \tag{A.58}$$

and so

$$\widehat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_i^0 = (\widehat{\mathbf{f}}^0' \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}^{0'} [-\mathbf{x}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) - \bar{\mathbf{e}}_r^0 \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i]. \tag{A.59}$$

From Proof of Theorem 1, we also have

$$\mathbf{x}_{i,t} - \mathbf{x}_i' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0' \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 = \boldsymbol{\tau}_{i,t} - \boldsymbol{\lambda}'_i \bar{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t} - (-\bar{\mathbf{e}}_r^0 \boldsymbol{\lambda}_i + \mathbf{v}_i)' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0' \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0. \tag{A.60}$$

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

$$\begin{aligned}
\widehat{\eta}_{i,t} &= \eta_{i,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' \mathbf{x}_{i,t} - \boldsymbol{\alpha}_i^{0'} (\widehat{\mathbf{f}}_t^0 - \mathbf{f}_t^0) - (\widehat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_i^0)' \widehat{\mathbf{f}}_t^0 + \boldsymbol{\varepsilon}_{i,t} \\
&= \eta_{i,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' \mathbf{x}_{i,t} - \boldsymbol{\alpha}_i' \bar{\mathbf{e}}_{r,t}^0 + \boldsymbol{\varepsilon}_{i,t} \\
&\quad - [-\mathbf{x}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) - \bar{\mathbf{e}}_r^0 \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i]' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0' \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 \\
&= \eta_{i,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' (\boldsymbol{\tau}_{i,t} - \boldsymbol{\lambda}'_i \bar{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t}) - \boldsymbol{\alpha}_i' \bar{\mathbf{e}}_{r,t}^0 + \boldsymbol{\varepsilon}_{i,t} \\
&\quad - [(-\bar{\mathbf{e}}_r^0 \boldsymbol{\lambda}_i + \mathbf{v}_i)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) - \bar{\mathbf{e}}_r^0 \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i]' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0' \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 \\
&= \eta_{i,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' \boldsymbol{\tau}_{i,t} - \tilde{\boldsymbol{\Lambda}}_{y,i}' \bar{\mathbf{e}}_{r,t}^0 - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' \mathbf{v}_{i,t} + \boldsymbol{\varepsilon}_{i,t} \\
&\quad - [-\bar{\mathbf{e}}_r^0 \tilde{\boldsymbol{\Lambda}}_{y,i} - \mathbf{v}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) + \boldsymbol{\varepsilon}_i]' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0' \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 \\
&= \eta_{i,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' \boldsymbol{\tau}_{i,t} - \tilde{\boldsymbol{\Lambda}}_{y,i}' \bar{\mathbf{e}}_{r,t}^0 - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' \mathbf{v}_{i,t} + \boldsymbol{\varepsilon}_{i,t} \\
&\quad - [-\bar{\mathbf{e}}_r^0 \tilde{\boldsymbol{\Lambda}}_{y,i} - \mathbf{v}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) + \boldsymbol{\varepsilon}_i]' \mathbf{f}^+ (\mathbf{f}^+)' (\mathbf{f}^+)'^{-1} \mathbf{f}_t^+ + O_p(N_0^{-1/2}) \\
&= \eta_{i,t} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' \boldsymbol{\tau}_{i,t} - \tilde{\boldsymbol{\Lambda}}_{y,i}' \bar{\mathbf{e}}_{r,t}^{0d} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i)' \mathbf{v}_{i,t}^d + \boldsymbol{\varepsilon}_{i,t}^d + O_p(N_0^{-1/2}),
\end{aligned} \tag{A.61}$$

where $\tilde{\boldsymbol{\Lambda}}_{y,i} := -\boldsymbol{\lambda}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_i) + \boldsymbol{\alpha}_i$.

Let us now consider $\eta_{i,t}$. We have assumed that $\Delta_{i,t} = \text{ATT}_{g,t} + v_{i,t}$, $\boldsymbol{\tau}_{i,t} = \boldsymbol{\tau}_{g,t} + \zeta_{i,t}$ and

$\beta_i = \beta + \nu_i$. This implies

$$\begin{aligned}
\eta_{i,t} &= \Delta_{i,t} - \tau'_{i,t} \beta_i \\
&= \text{ATT}_{g,t} + v_{i,t} - (\tau_{g,t} + \zeta_{i,t})'(\beta + \nu_i) \\
&= \text{ATT}_{g,t} - \tau'_{g,t} \beta + v_{i,t} - \tau'_{g,t} \nu_i - \zeta'_{i,t}(\beta + \nu_i) \\
&= \eta_{g,t} + \epsilon_{i,t},
\end{aligned} \tag{A.62}$$

where $\epsilon_{i,t} := v_{i,t} - \tau'_{g,t} \nu_i - \zeta'_{i,t} \beta_i$ has the same basic properties as $v_{i,t}$. By using this, the above expression for $\hat{\eta}_{i,t}$ and the arguments provided in Proof of Theorem 1 for why averaging and scaling by $\sqrt{N_g}$ does not affect the order of the reminder, we obtain

$$\begin{aligned}
&\sqrt{N_g}(\widehat{\text{DATT}}_{g,t} - \text{DATT}_{g,t}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (\hat{\eta}_{i,t} - \text{DATT}_{g,t}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (\hat{\eta}_{i,t} - \eta_{i,t} + \epsilon_{i,t}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} [\epsilon_{i,t} - (\hat{\beta} - \beta_i)' \tau_{i,t} - \tilde{\Lambda}'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d} - (\hat{\beta} - \beta_i)' \mathbf{v}_{i,t}^d + \epsilon_{i,t}^d] + O_p(N_0^{-1/2}). \tag{A.63}
\end{aligned}$$

Consider the second term on the right. From $\beta_i = \beta + \nu_i$,

$$\frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (\hat{\beta} - \beta_i)' \tau_{i,t} = \sqrt{\frac{N_g}{N}} \sqrt{N} (\hat{\beta} - \beta)' \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \tau_{i,t} - \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} \nu'_i \tau_{i,t}. \tag{A.64}$$

By using the same steps as in the proof of Theorem A.1 below, we can show that $\sqrt{N}(\hat{\beta} - \beta) = O_p(1)$. Hence, if we in addition assume as in Theorem 2 that $N_g/N_0 \rightarrow 0$, then we also have $N_g/N \rightarrow 0$, as $N > N_0$, which in turn implies

$$\frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (\hat{\beta} - \beta_i)' \tau_{i,t} = -\frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} \nu'_i \tau_{i,t} + o_p(1). \tag{A.65}$$

We can similarly show that the error coming from the estimation of β is negligible also in the term involving $(\hat{\beta} - \beta_i)' \mathbf{v}_{i,t}^d$. The term that involves $\tilde{\Lambda}'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d}$ is negligible as a whole, as is apparent from

$$\frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} \tilde{\Lambda}'_{y,i} \bar{\mathbf{e}}_{r,t}^{0d} = \sqrt{\frac{N_g}{N_0}} \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \tilde{\Lambda}'_{y,i} \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} = o_p(1), \tag{A.66}$$

which holds since again $N_g/N_0 \rightarrow 0$. By adding these results, we arrive at the following limiting expression for $\sqrt{N_g}(\widehat{\text{DATT}}_{g,t} - \text{DATT}_{g,t})$:

$$\begin{aligned}\sqrt{N_g}(\widehat{\text{DATT}}_{g,t} - \text{DATT}_{g,t}) &= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (\epsilon_{i,t} + \nu_i' \tau_{i,t} + \nu_i' \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) + o_p(1) \\ &= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} - \zeta_{i,t}' \beta + \nu_i' \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) + o_p(1),\end{aligned}\quad (\text{A.67})$$

where the last equality holds, because $\epsilon_{i,t} + \nu_i' \tau_{i,t} = v_{i,t} - \tau_{g,t}' \nu_i - \zeta_{i,t}' \beta + \nu_i' \tau_{i,t} = v_{i,t} - \zeta_{i,t}' \beta$, and where the first term on the right-hand side of the same equality is asymptotically mixed normal by the same arguments used in Proof of Theorem 1.

Let us now consider the estimator of the IATT, $\widehat{\text{IATT}}_{g,t} := \widehat{\text{ATT}}_{g,t} - \widehat{\text{DATT}}_{g,t}$. By the definition of $\text{DATT}_{g,t}$, the above expression for $\sqrt{N_g}(\widehat{\text{DATT}}_{g,t} - \text{DATT}_{g,t})$ and the one given for $\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t})$ in Proof of Theorem 1,

$$\begin{aligned}\sqrt{N_g}(\widehat{\text{IATT}}_{g,t} - \tau_{g,t}' \beta) &= \sqrt{N_g}[\widehat{\text{ATT}}_{g,t} - \widehat{\text{DATT}}_{g,t} - (\text{ATT}_{g,t} - \text{DATT}_{g,t})] \\ &= \sqrt{N_g}[\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t} - (\widehat{\text{DATT}}_{g,t} - \text{DATT}_{g,t})] \\ &= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} [v_{i,t} + \beta_i' \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d - (v_{i,t} - \zeta_{i,t}' \beta + \nu_i' \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d)] + o_p(1) \\ &= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (\zeta_{i,t} + \mathbf{v}_{i,t}^d)' \beta + o_p(1),\end{aligned}\quad (\text{A.68})$$

where the first term on the right is again asymptotically mixed normal. ■

2 Covariates without common factor representation

In the main paper, we assumed that all the regressors admitted to the common factor representation in Assumption 4, which rules out specifications featuring, for example, dummy variables, powers or products of the regressors. In the present section, we show that this assumption is not necessary, and that it can be relaxed provided that the estimation procedure outlined in the main paper are suitably modified.

The model we have in mind for $y_{i,t}$ for the untreated units is now given by

$$y_{i,t}(0) = \theta' \mathbf{w}_{i,t} + \beta_i' \mathbf{x}_{i,t}(0) + \alpha_i' \mathbf{f}_t + \varepsilon_{i,t}, \quad (\text{A.69})$$

where $\mathbf{w}_{i,t}$ is an $n \times 1$ vector of regressors that unlike $\mathbf{x}_{i,t}(0)$ do not have a common factor representation. The new regressors can be correlated with \mathbf{f}_t ; however, we do require that they are independent of $\varepsilon_{i,t}$ and $\mathbf{v}_{i,t}$. They are in this sense “conditionally exogenous” (given \mathbf{f}_t).¹ We assume that the coefficients of $\mathbf{w}_{i,t}$ are equal across the cross-section. This is not necessary; however, it simplifies the derivations. At the end of this section, we explain how the equal slope condition can be relaxed and what the consequences are.

The presence of $\mathbf{w}_{i,t}$ requires two changes to the counterfactual estimation procedure outlined in the main paper;

1. The step-1 estimate of the unknown factors is no longer given by $\widehat{\mathbf{f}}_t = \bar{\mathbf{z}}_t$ but is instead given by

$$\widehat{\mathbf{f}}_t = \begin{bmatrix} \bar{\mathbf{z}}_t \\ \mathbf{w}_t \end{bmatrix}, \quad (\text{A.70})$$

where $\bar{\mathbf{a}}_t := N_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbf{a}_{i,t}$ is the cross-sectional average of any vector $\mathbf{a}_{i,t}$ for the group of untreated units ($g = 0$), just as before.

2. The step-3 counterfactual estimator should be changed from $\widehat{y}_{i,t}(0) = \widehat{\mathbf{a}}_i' \widehat{\mathbf{f}}_t$ to

$$\widehat{y}_{i,t}(0) = \widehat{\boldsymbol{\theta}}' \mathbf{w}_{i,t} + \widehat{\mathbf{a}}_i' \widehat{\mathbf{f}}_t, \quad (\text{A.71})$$

where

$$\widehat{\mathbf{a}}_i = (\widehat{\mathbf{f}}' \widehat{\mathbf{f}})^{-1} \widehat{\mathbf{f}}' (\mathbf{y}_i - \mathbf{w}_i \widehat{\boldsymbol{\theta}}), \quad (\text{A.72})$$

$$\widehat{\boldsymbol{\theta}} = \left(\sum_{i=1}^N \mathbf{w}'_i \mathbf{M}_{\widehat{\mathbf{f}}} \mathbf{w}_i \right)^{-1} \sum_{i=1}^N \mathbf{w}'_i \mathbf{M}_{\widehat{\mathbf{f}}} \mathbf{y}_i, \quad (\text{A.73})$$

where \mathbf{w}_i , $\widehat{\mathbf{f}}$ and \mathbf{y}_i are T_0 -rowed matrices of stacked observations on $\mathbf{w}_{i,t}$, $\widehat{\mathbf{f}}_t$ and $y_{i,t}$, respectively.

With the above changes in place, the appropriate ATT estimator for group g at time t is given by

$$\widehat{\text{ATT}}_{g,t} = \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \widehat{\Delta}_{i,t} = \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} [y_{i,t} - \widehat{y}_{i,t}(0)]. \quad (\text{A.74})$$

¹Conditional exogeneity rules out predetermined covariates like lagged dependent variables. This can be relaxed but then $\varepsilon_{i,t}$ can no longer be serially correlated.

An important property of the counterfactual estimation procedure is that it is exactly invariant with respect to augmenting $\mathbf{w}_{i,t}$ in step 2 by $\mathbf{x}_{i,t}$. This is very important because it means that researchers do not have to know which regressors that admit to a common factor representation and which ones that do not; that is, researchers do not have to be able to distinguish between $\mathbf{x}_{i,t}$ and $\mathbf{w}_{i,t}$.

Theorem A.1 provides the asymptotic distribution of the above version of $\widehat{\text{ATT}}_{g,t}$. De Vos and Westerlund (2019) consider regular CCE estimation in the presence of regressors that do not load on the same set of factors as the dependent variable. However, they do not provide any formal analysis but just a heuristic discussion. As far as we are aware, Theorem A.1 is the first formal result that allow for conditionally exogenous regressors in the CCE context.

Theorem A.1. *Suppose that Assumptions 1–10 and the above stated conditions are met. Then, as $N_g, N_0 \rightarrow \infty$ with $N_g/N_0 \rightarrow 0$,*

$$(a) \frac{\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t})}{\sqrt{\text{var}(\widehat{\text{ATT}}_{g,t})}} \rightarrow_d N(0, 1),$$

$$(b) \widehat{\text{var}}(\widehat{\text{ATT}}_{g,t}) \rightarrow_p \text{var}(\widehat{\text{ATT}}_{g,t}),$$

where $\text{var}(\widehat{\text{ATT}}_{g,t}) = \text{var}[\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t})|\mathcal{W}]$ and \mathcal{W} is the sigma-field generated by $(\mathbf{f}, \bar{\mathbf{w}})$.

Proof: We begin this proof by studying the effect of the new regressors on the estimated factors. Since $y_{i,t} = \boldsymbol{\theta}'\mathbf{w}_{i,t} + \boldsymbol{\beta}'_i\mathbf{x}_{i,t} + \boldsymbol{\alpha}'_i\mathbf{f}_t + \varepsilon_{i,t}$ for the group of untreated units ($i \in \mathcal{I}_0$), we have

$$\bar{\mathbf{z}}_t = \bar{\Lambda}'\mathbf{f}_t + \mathbb{1}_1\boldsymbol{\theta}'\bar{\mathbf{w}}_t + \bar{\mathbf{e}}_t, \quad (\text{A.75})$$

where $\mathbb{1}_1 := [1, 0, \dots, 0]'$ is a $(m+1) \times 1$ vector. This implies the following representation for $\widehat{\mathbf{f}}_t$:

$$\begin{aligned} \widehat{\mathbf{f}}_t &= \begin{bmatrix} \bar{\mathbf{z}}_t \\ \bar{\mathbf{w}}_t \end{bmatrix} \\ &= \begin{bmatrix} \bar{\Lambda}'\mathbf{f}_t + \mathbb{1}_1\boldsymbol{\theta}'\bar{\mathbf{w}}_t + \bar{\mathbf{e}}_t \\ \bar{\mathbf{w}}_t \end{bmatrix} \\ &= \begin{bmatrix} \bar{\Lambda}' & \mathbb{1}_1\boldsymbol{\theta}' \\ \mathbf{0}_{n \times r} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{f}_t \\ \bar{\mathbf{w}}_t \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{e}}_t \\ \mathbf{0}_{n \times 1} \end{bmatrix} \\ &= \bar{\Lambda}'_*\mathbf{f}_{*,t} + \bar{\mathbf{e}}_{*,t}, \end{aligned} \quad (\text{A.76})$$

with obvious definitions of $\bar{\Lambda}_*$, $\mathbf{f}_{*,t}$ and $\bar{\mathbf{e}}_{*,t}$. Hence, defining

$$\bar{\mathbf{H}}_* := \begin{bmatrix} \bar{\mathbf{H}} & \mathbf{0}_{(m+1) \times n} \\ \mathbf{0}'_{(m+1) \times n} & \mathbf{I}_n \end{bmatrix}, \quad (\text{A.77})$$

an $(m+1+n) \times (m+1+n)$ matrix, we get

$$\hat{\mathbf{f}}_t^0 = \bar{\mathbf{H}}_*' \hat{\mathbf{f}}_t = \bar{\mathbf{H}}_*' \bar{\Lambda}'_* \mathbf{f}_{*,t} + \bar{\mathbf{H}}_*' \bar{\mathbf{e}}_{*,t} = \mathbf{f}_{*,t}^0 + \bar{\mathbf{e}}_{*,t}^0 \quad (\text{A.78})$$

with implicit definitions of $\mathbf{f}_{*,t}^0$ and $\bar{\mathbf{e}}_{*,t}^0$. Note how $\bar{\mathbf{e}}_{*,t}^0 = [\bar{\mathbf{e}}_t' \bar{\mathbf{H}}, \mathbf{0}'_{n \times 1}]'$. Further use of $\bar{\mathbf{H}}' \bar{\Lambda}' = [\mathbf{I}_r, \mathbf{0}'_{(m+1-r) \times r}]'$ and $\bar{\mathbf{H}} = [\bar{\mathbf{H}}_r, \bar{\mathbf{H}}_{-r}]$, reveals that $\mathbf{f}_{*,t}^0$ has the following form:

$$\begin{aligned} \mathbf{f}_{*,t}^0 &= \bar{\mathbf{H}}_*' \bar{\Lambda}'_* \mathbf{f}_{*,t} \\ &= \begin{bmatrix} \bar{\mathbf{H}}' \bar{\Lambda}' & \bar{\mathbf{H}}' \mathbb{1}_1 \theta' \\ \mathbf{0}_{n \times r} & \mathbf{I}_n \end{bmatrix} \mathbf{f}_{*,t} \\ &= \begin{bmatrix} \bar{\mathbf{H}}' \bar{\Lambda}' \mathbf{f}_t + \bar{\mathbf{H}}' \mathbb{1}_1 \theta' \bar{\mathbf{w}}_t \\ \bar{\mathbf{w}}_t \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{f}_t + \bar{\mathbf{H}}_r' \mathbb{1}_1 \theta' \bar{\mathbf{w}}_t \\ \bar{\mathbf{H}}_{-r}' \mathbb{1}_1 \theta' \bar{\mathbf{w}}_t \\ \bar{\mathbf{w}}_t \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{f}_t \\ \mathbf{0}_{(m+1+n-r) \times 1} \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{H}}' \mathbb{1}_1 \theta' \\ \mathbf{I}_n \end{bmatrix} \bar{\mathbf{w}}_t. \end{aligned} \quad (\text{A.79})$$

The rotated factors in $\mathbf{f}_{*,t}^0$ are the objects that will be estimated. Suppose that the rows of $\bar{\mathbf{w}}_t$ are all different from zero. Then the above equation reveals that, unlike \mathbf{f}_t^0 , all the rows of $\mathbf{f}_{*,t}^0$ are nonzero. Hence, unlike in Proof of Theorem 1, here there is no need to rescale by \mathbf{D}_{N_0} .

We now turn our attention to $\hat{\Delta}_{i,t}$. From $y_{i,t} = \eta_{i,t} + \theta' \mathbf{w}_{i,t} + \beta'_i \mathbf{x}_{i,t} + \alpha'_i \mathbf{f}_t + \varepsilon_{i,t}$ and $\hat{y}_{i,t}(0) = \hat{\theta}' \mathbf{w}_{i,t} + \hat{\alpha}'_i \hat{\mathbf{f}}_t$,

$$\begin{aligned} \hat{\Delta}_{i,t} &= y_{i,t} - \hat{y}_{i,t}(0) \\ &= \eta_{i,t} + \theta' \mathbf{w}_{i,t} + \beta'_i \mathbf{x}_{i,t} + \alpha'_i \mathbf{f}_t + \varepsilon_{i,t} - (\hat{\theta}' \mathbf{w}_{i,t} + \hat{\alpha}'_i \hat{\mathbf{f}}_t) \\ &= \eta_{i,t} - (\hat{\theta} - \theta)' \mathbf{w}_{i,t} + \beta'_i \mathbf{x}_{i,t} - (\hat{\alpha}'_i \hat{\mathbf{f}}_t - \alpha'_i \mathbf{f}_t) + \varepsilon_{i,t} \\ &= \eta_{i,t} - (\hat{\theta} - \theta)' \mathbf{w}_{i,t} + \beta'_i \mathbf{x}_{i,t} - \alpha'^0_i (\hat{\mathbf{f}}_t^0 - \mathbf{f}_{*,t}^0) - (\hat{\alpha}_i^0 - \alpha_{*,i}^0)' \hat{\mathbf{f}}_t^0 + \varepsilon_{i,t}. \end{aligned} \quad (\text{A.80})$$

For the last equality we use the fact that the rank of the $(m+1+n) \times (r+n)$ matrix $\bar{\Lambda}'_*$ is given by $\text{rk } \bar{\Lambda}'_* + \text{rk } \mathbf{I}_n = r+n$ (see Abadir and Magnus, 2005, Exercise 5.48). It therefore has

full column rank, which in view of $\text{rk}(\bar{\mathbf{H}}'_*\bar{\boldsymbol{\Lambda}}'_*) = \text{rk} \bar{\boldsymbol{\Lambda}}'_*$ (see Abadir and Magnus, 2005, Exercise 4.24) in turn implies that $\bar{\mathbf{H}}'_*\bar{\boldsymbol{\Lambda}}'_*$ has full column rank, too. This means that its Moore–Penrose inverse is given by $(\bar{\mathbf{H}}'_*\bar{\boldsymbol{\Lambda}}'_*)^+ = (\bar{\boldsymbol{\Lambda}}_*\bar{\mathbf{H}}_*\bar{\mathbf{H}}'_*\bar{\boldsymbol{\Lambda}}'_*)^{-1}\bar{\boldsymbol{\Lambda}}_*\bar{\mathbf{H}}_*$ and hence that $(\bar{\mathbf{H}}'_*\bar{\boldsymbol{\Lambda}}'_*)^+\bar{\mathbf{H}}'_*\bar{\boldsymbol{\Lambda}}'_* = \mathbf{I}_{r+n}$. We also note that the $(m+1+n) \times (m+1+n)$ matrix $\bar{\mathbf{H}}_*$ is block diagonal with positive definite blocks. It is therefore invertible. Hence, letting $\hat{\mathbf{a}}_i^0 := (\bar{\mathbf{H}}'_*)^{-1}\hat{\mathbf{a}}_i = \bar{\mathbf{H}}_*^{-1}\hat{\mathbf{a}}_i$ and $\boldsymbol{\alpha}_i^0 := (\bar{\mathbf{H}}'_*\bar{\boldsymbol{\Lambda}}'_*)^{+\prime}\boldsymbol{\alpha}_{*,i}$, where $\boldsymbol{\alpha}_{*,i} := [\boldsymbol{\alpha}'_i, \mathbf{0}'_{n \times 1}]'$, we obtain

$$\begin{aligned}\hat{\mathbf{a}}_i'\hat{\mathbf{f}}_t - \boldsymbol{\alpha}_i'\mathbf{f}_t &= \hat{\mathbf{a}}_i'\hat{\mathbf{f}}_t - [\boldsymbol{\alpha}'_i, \mathbf{0}'_{n \times 1}] \begin{bmatrix} \mathbf{f}_t \\ \bar{\mathbf{w}}_t \end{bmatrix} \\ &= \hat{\mathbf{a}}_i'(\bar{\mathbf{H}}'_*)^{-1}\bar{\mathbf{H}}'_*\hat{\mathbf{f}}_t - \boldsymbol{\alpha}'_{*,i}(\bar{\mathbf{H}}'_*\bar{\boldsymbol{\Lambda}}'_*)^+\bar{\mathbf{H}}'_*\bar{\boldsymbol{\Lambda}}'_*\mathbf{f}_{*,t} \\ &= \hat{\mathbf{a}}_i'^0\hat{\mathbf{f}}_t^0 - \boldsymbol{\alpha}_{*,i}^{0\prime}\mathbf{f}_{*,t}^0 \\ &= \boldsymbol{\alpha}_{*,i}^{0\prime}(\hat{\mathbf{f}}_t^0 - \mathbf{f}_{*,t}^0) + (\hat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_{*,i}^0)' \hat{\mathbf{f}}_t^0.\end{aligned}\tag{A.81}$$

Consider $\hat{\mathbf{a}}_i^0 - \boldsymbol{\alpha}_{*,i}^0$. From $\bar{\boldsymbol{\Lambda}}_*\bar{\boldsymbol{\Lambda}}_*^+ = \mathbf{I}_{r+n}$, letting $\mathbf{a}_{*,i} := \bar{\boldsymbol{\Lambda}}_*^+\boldsymbol{\alpha}_{*,i}$,

$$\begin{aligned}\mathbf{y}_i &= \mathbf{w}_i\theta + \mathbf{x}_i\beta_i + \mathbf{f}\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i \\ &= \mathbf{w}_i\theta + \mathbf{x}_i\beta_i + \mathbf{f}_*\bar{\boldsymbol{\Lambda}}_*\bar{\boldsymbol{\Lambda}}_*^+\boldsymbol{\alpha}_{*,i} + \boldsymbol{\varepsilon}_i \\ &= \mathbf{w}_i\theta + \mathbf{x}_i\beta_i + \hat{\mathbf{f}}\bar{\boldsymbol{\Lambda}}_*^+\boldsymbol{\alpha}_{*,i} - (\hat{\mathbf{f}} - \mathbf{f}_*\bar{\boldsymbol{\Lambda}}_*)\bar{\boldsymbol{\Lambda}}_*^+\boldsymbol{\alpha}_{*,i} + \boldsymbol{\varepsilon}_i \\ &= \mathbf{w}_i\theta + \mathbf{x}_i\beta_i + \hat{\mathbf{f}}\mathbf{a}_{*,i} - \bar{\mathbf{e}}_*\mathbf{a}_{*,i} + \boldsymbol{\varepsilon}_i.\end{aligned}\tag{A.82}$$

We can now use the same steps as in the proof of Theorem 1 to show that

$$\begin{aligned}\hat{\mathbf{a}}_i^0 &= \bar{\mathbf{H}}_*^{-1}\hat{\mathbf{a}}_i \\ &= \bar{\mathbf{H}}_*^{-1}(\hat{\mathbf{f}}\hat{\mathbf{f}}')^{-1}\hat{\mathbf{f}}'(\mathbf{y}_i - \mathbf{w}_i\hat{\theta}) \\ &= \bar{\mathbf{H}}_*^{-1}\mathbf{a}_{*,i} + \bar{\mathbf{H}}_*^{-1}(\hat{\mathbf{f}}\hat{\mathbf{f}}')^{-1}\hat{\mathbf{f}}'[-\mathbf{w}_i(\hat{\theta} - \theta) + \mathbf{x}_i\beta_i - \bar{\mathbf{e}}_*\mathbf{a}_{*,i} + \boldsymbol{\varepsilon}_i] \\ &= \bar{\mathbf{H}}_*^{-1}\mathbf{a}_{*,i} + (\bar{\mathbf{H}}'_*\hat{\mathbf{f}}'\hat{\mathbf{f}}\bar{\mathbf{H}}_*)^{-1}\bar{\mathbf{H}}'_*\hat{\mathbf{f}}'[-\mathbf{w}_i(\hat{\theta} - \theta) + \mathbf{x}_i\beta_i - \bar{\mathbf{e}}_*\mathbf{a}_{*,i} + \boldsymbol{\varepsilon}_i] \\ &= \boldsymbol{\alpha}_{*,i}^0 + (\hat{\mathbf{f}}^0\hat{\mathbf{f}}^0)^{-1}\hat{\mathbf{f}}^{0\prime}[-\mathbf{w}_i(\hat{\theta} - \theta) + \mathbf{x}_i\beta_i - \bar{\mathbf{e}}_*\mathbf{a}_{*,i} + \boldsymbol{\varepsilon}_i]\end{aligned}\tag{A.83}$$

where the last equality holds because $\bar{\mathbf{H}}_*^{-1}\mathbf{a}_{*,i} = \bar{\mathbf{H}}_*^{-1}\bar{\boldsymbol{\Lambda}}_*^+\boldsymbol{\alpha}_{*,i} = (\bar{\boldsymbol{\Lambda}}_*\bar{\mathbf{H}}_*)^+\boldsymbol{\alpha}_{*,i} = \boldsymbol{\alpha}_{*,i}^0$. Note also how

$$\boldsymbol{\alpha}_{*,i}^{0\prime}(\hat{\mathbf{f}}_t^0 - \mathbf{f}_{*,t}^0) = \boldsymbol{\alpha}_{*,i}^{0\prime}\bar{\mathbf{e}}_{*,t}^0 = \boldsymbol{\alpha}_{*,i}'(\bar{\mathbf{H}}'_*\bar{\boldsymbol{\Lambda}}'_*)^+\bar{\mathbf{H}}'_*\bar{\mathbf{e}}_{*,t} = \boldsymbol{\alpha}_{*,i}'\bar{\boldsymbol{\Lambda}}_*^{+\prime}\bar{\mathbf{e}}_{*,t} = \mathbf{a}_{*,i}'\bar{\mathbf{e}}_{*,t}.\tag{A.84}$$

Direct insertion into the above expression for $\widehat{\Delta}_{i,t}$ yields

$$\begin{aligned}\widehat{\Delta}_{i,t} &= \eta_{i,t} - (\widehat{\theta} - \theta)' \mathbf{w}_{i,t} + \beta_i' \mathbf{x}_{i,t} - \alpha_{*,i}' (\widehat{\mathbf{f}}_t^0 - \mathbf{f}_{*,t}^0) - (\widehat{\mathbf{a}}_i^0 - \alpha_{*,i}^0)' \widehat{\mathbf{f}}_t^0 + \varepsilon_{i,t} \\ &= \eta_{i,t} - (\widehat{\theta} - \theta)' \mathbf{w}_{i,t} + \beta_i' \mathbf{x}_{i,t} - \mathbf{a}_{*,i}' \bar{\mathbf{e}}_{*,t} + \varepsilon_{i,t} \\ &\quad - [-\mathbf{w}_i(\widehat{\theta} - \theta) + \mathbf{x}_i \beta_i - \bar{\mathbf{e}}_* \mathbf{a}_{*,i} + \varepsilon_i] \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0 \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0.\end{aligned}\tag{A.85}$$

Let us define $\lambda_{*,i} := [\lambda_i', \mathbf{0}_{n \times 1}]'$ analogously to $\alpha_{*,i}$ and $\mathbf{l}_{*,i} := \bar{\Lambda}_*^+ \lambda_{*,i}$ analogously to $\mathbf{a}_{*,i}$. Then, via $\bar{\Lambda}_* \bar{\Lambda}_*^+ = \mathbf{I}_{r+n}$ and $\widehat{\mathbf{f}} = \widehat{\mathbf{f}}^0 \bar{\mathbf{H}}_*^{-1}$,

$$\begin{aligned}\mathbf{x}_i &= \mathbf{f} \lambda_i + \mathbf{v}_i \\ &= \mathbf{f}_* \lambda_{*,i} + \mathbf{v}_i \\ &= \widehat{\mathbf{f}} \bar{\Lambda}_*^+ \lambda_{*,i} - (\widehat{\mathbf{f}} - \mathbf{f}_* \bar{\Lambda}_*) \bar{\Lambda}_*^+ \lambda_{*,i} + \mathbf{v}_i \\ &= \widehat{\mathbf{f}}^0 \bar{\mathbf{H}}_*^{-1} \bar{\Lambda}_*^+ \lambda_{*,i} - \bar{\mathbf{e}}_* \bar{\Lambda}_*^+ \lambda_{*,i} + \mathbf{v}_i \\ &= \widehat{\mathbf{f}}^0 \bar{\mathbf{H}}_*^{-1} \mathbf{l}_{*,i} - \bar{\mathbf{e}}_* \mathbf{l}_{*,i} + \mathbf{v}_i\end{aligned}\tag{A.86}$$

if $t \leq T_0$, whereas if $t > T_0$, then

$$\mathbf{x}_{i,t} = \tau_{i,t} + \lambda_{*,i}' \mathbf{f}_{*,t} + \mathbf{v}_{i,t} = \tau_{i,t} + \mathbf{l}_{*,i}' \bar{\mathbf{H}}_*^{-1} \widehat{\mathbf{f}}_t^0 - \mathbf{l}_{*,i}' \bar{\mathbf{e}}_{*,t} + \mathbf{v}_{i,t}.\tag{A.87}$$

Hence, similarly to Proof of Theorem 1,

$$\begin{aligned}\widehat{\Delta}_{i,t} &= \Delta_{i,t} - (\widehat{\theta} - \theta)' \mathbf{w}_{i,t} + \beta_i' (-\mathbf{l}_{*,i}' \bar{\mathbf{e}}_{*,t} + \mathbf{v}_{i,t}) - \mathbf{a}_{*,i}' \bar{\mathbf{e}}_{*,t} + \varepsilon_{i,t} \\ &\quad - [-\mathbf{w}_i(\widehat{\theta} - \theta) + (-\bar{\mathbf{e}}_* \mathbf{l}_{*,i} + \mathbf{v}_i) \beta_i - \bar{\mathbf{e}}_* \mathbf{a}_{*,i} + \varepsilon_i] \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0 \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 \\ &= \Delta_{i,t} - (\widehat{\theta} - \theta)' \mathbf{w}_{i,t} - \Lambda_{*,y,i}' \bar{\mathbf{e}}_{*,t} + \beta_i' \mathbf{v}_{i,t} + \varepsilon_{i,t} \\ &\quad - [-\mathbf{w}_i(\widehat{\theta} - \theta) - \bar{\mathbf{e}}_* \Lambda_{*,y,i} + \mathbf{v}_i \beta_i + \varepsilon_i] \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0 \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0.\end{aligned}\tag{A.88}$$

where we have defined $\Lambda_{*,y,i} := \mathbf{l}_{*,i} \beta_i + \mathbf{a}_{*,i}$ similarly to $\Lambda_{y,i}$

The above expression is entirely analogous to the one derived in Proof of Theorem 1. The main difference is the terms involving $\widehat{\theta} - \theta$. Let us therefore now consider this quantity. From $\mathbf{y}_i = \mathbf{w}_i \theta + \mathbf{x}_i \beta_i + \widehat{\mathbf{f}} \mathbf{a}_{*,i} - \bar{\mathbf{e}}_* \mathbf{a}_{*,i} + \varepsilon_i$, $\mathbf{x}_i = \widehat{\mathbf{f}}^0 \bar{\mathbf{H}}_*^{-1} \mathbf{l}_{*,i} - \bar{\mathbf{e}}_* \mathbf{l}_{*,i} + \mathbf{v}_i$ and $\mathbf{M}_{\widehat{\mathbf{f}}} = \mathbf{M}_{\widehat{\mathbf{f}}^0}$,

$$\begin{aligned}\widehat{\theta} - \theta &= \left(\sum_{i=1}^N \mathbf{w}_i' \mathbf{M}_{\widehat{\mathbf{f}}} \mathbf{w}_i \right)^{-1} \sum_{i=1}^N \mathbf{w}_i' \mathbf{M}_{\widehat{\mathbf{f}}} (\mathbf{x}_i \beta_i - \bar{\mathbf{e}}_* \mathbf{a}_{*,i} + \varepsilon_i) \\ &= \left(\sum_{i=1}^N \mathbf{w}_i' \mathbf{M}_{\widehat{\mathbf{f}}^0} \mathbf{w}_i \right)^{-1} \sum_{i=1}^N \mathbf{w}_i' \mathbf{M}_{\widehat{\mathbf{f}}^0} (-\bar{\mathbf{e}}_* \Lambda_{*,y,i} + \mathbf{v}_i \beta_i + \varepsilon_i).\end{aligned}\tag{A.89}$$

Consider the denominator. Making use of (A.12) in Westerlund et al. (2019), and the fact that $\|\bar{\mathbf{e}}_*^0\|$ and $\|(\hat{\mathbf{f}}^0 \hat{\mathbf{f}}^0)^+ - (\mathbf{f}_*^0 \mathbf{f}_*^0)^+\|$ are both $O_p(N_0^{-1/2})$ by the same arguments as in Proof of Theorem 1, we have

$$\begin{aligned}
& \|\mathbf{M}_{\mathbf{f}^0} - \mathbf{M}_{\hat{\mathbf{f}}_*^0}\| \\
&= \|\bar{\mathbf{e}}_*^0(\hat{\mathbf{f}}^0 \hat{\mathbf{f}}^0)^+ \bar{\mathbf{e}}_*^{0'} + \bar{\mathbf{e}}_*^0(\hat{\mathbf{f}}^0 \hat{\mathbf{f}}^0)^+ \mathbf{f}^{0'} + \mathbf{f}^0(\hat{\mathbf{f}}^0 \hat{\mathbf{f}}^0)^+ \bar{\mathbf{e}}_*^{0'} + \mathbf{f}_*^0[(\hat{\mathbf{f}}^0 \hat{\mathbf{f}}^0)^+ - (\mathbf{f}_*^0 \mathbf{f}_*^0)^+] \mathbf{f}_*^{0'}\| \\
&\leq \|\bar{\mathbf{e}}_*^0\|^2 \|(\hat{\mathbf{f}}^0 \hat{\mathbf{f}}^0)^+\| + 2\|\bar{\mathbf{e}}_*^0\| \|(\hat{\mathbf{f}}^0 \hat{\mathbf{f}}^0)^+\| \|\mathbf{f}_*^0\| + \|\mathbf{f}_*^0\|^2 \|(\hat{\mathbf{f}}^0 \hat{\mathbf{f}}^0)^+ - (\mathbf{f}_*^0 \mathbf{f}_*^0)^+\| \\
&= O_p(N_0^{-1/2})
\end{aligned} \tag{A.90}$$

which in turn implies

$$\left\| \frac{1}{N} \sum_{i=1}^N \mathbf{w}'_i (\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}) \mathbf{w}_i \right\| \leq \frac{1}{N} \sum_{i=1}^N \|\mathbf{w}_i\|^2 \|\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}\| = O_p(N_0^{-1/2}). \tag{A.91}$$

It follows that

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^N \mathbf{w}'_i \mathbf{M}_{\hat{\mathbf{f}}^0} \mathbf{w}_i &= \frac{1}{N} \sum_{i=1}^N \mathbf{w}'_i \mathbf{M}_{\mathbf{f}_*^0} \mathbf{w}_i - \frac{1}{N} \sum_{i=1}^N \mathbf{w}'_i (\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}) \mathbf{w}_i \\
&= \frac{1}{N} \sum_{i=1}^N \mathbf{w}'_i \mathbf{M}_{\mathbf{f}_*^0} \mathbf{w}_i + O_p(N_0^{-1/2}) \\
&= \Sigma_{\mathbf{w}^d} + o_p(1),
\end{aligned} \tag{A.92}$$

where $\Sigma_{\mathbf{w}^d} := \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \mathbb{E}(\mathbf{w}'_i \mathbf{M}_{\mathbf{f}_*^0} \mathbf{w}_i | \mathcal{W}) = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \mathbb{E}(\mathbf{w}_i^{d'} \mathbf{w}_i^d | \mathcal{W})$ with $\mathbf{w}_i^d = \mathbf{M}_{\mathbf{f}_*^0} \mathbf{w}_i$ being the stacked, defactored \mathbf{w}_i , similarly to Proof of Theorem 1. Also, \mathcal{W} is the sigma-field generated by \mathbf{f}_*^0 . We assume that $\Sigma_{\mathbf{w}^d}$ is positive definite.

As for the numerator,

$$\begin{aligned}
& \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{w}'_i (\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}) \bar{\mathbf{e}}_*^0 \Lambda_{*,y,i} \right\| \\
&\leq \frac{\sqrt{N}}{N_0} \frac{1}{N} \sum_{i=1}^N \|\mathbf{w}_i\| \|\Lambda_{*,y,i}\| \sqrt{N_0} \|\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}\| \sqrt{N_0} \|\bar{\mathbf{e}}_*^0\| = O_p(\sqrt{NN_0^{-1}})
\end{aligned} \tag{A.93}$$

By Exercise 10.20 in Abadir and Magnus (2005), $\|\text{vec } \mathbf{A}\|^2 = (\text{vec } \mathbf{A})' \text{vec } \mathbf{A} = \text{tr}(\mathbf{A}' \mathbf{A}) = \|\mathbf{A}\|^2$. By using this and $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \text{vec } \mathbf{B}$ (see Abadir and Magnus, 2005, Exercise

10.18), we obtain

$$\begin{aligned}
\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{w}'_i (\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}) \boldsymbol{\varepsilon}_i \right\| &= \left\| \text{vec} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{w}'_i (\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}) \boldsymbol{\varepsilon}_i \right) \right\| \\
&= \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N (\boldsymbol{\varepsilon}'_i \otimes \mathbf{w}'_i) \text{vec} (\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}) \right\| \\
&= \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N (\boldsymbol{\varepsilon}_i \otimes \mathbf{w}_i)' \text{vec} (\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}) \right\| \\
&\leq \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N (\boldsymbol{\varepsilon}_i \otimes \mathbf{w}_i) \right\| \|\text{vec} (\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0})\| \\
&= \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N (\boldsymbol{\varepsilon}_i \otimes \mathbf{w}_i) \right\| \|\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}\| \\
&= O_p(N_0^{-1/2}),
\end{aligned} \tag{A.94}$$

where the last equality supposes that \mathbf{w}_i is independent of $\boldsymbol{\varepsilon}_i$. The order of $N^{-1/2} \sum_{i=1}^N \mathbf{w}'_i (\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}) \mathbf{v}_i \boldsymbol{\beta}_i$ is the same, provided that \mathbf{w}_i is independent also of \mathbf{v}_i . Hence, by adding the results,

$$\begin{aligned}
&\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{w}'_i \mathbf{M}_{\hat{\mathbf{f}}^0} (-\bar{\mathbf{e}}_* \boldsymbol{\Lambda}_{*,y,i} + \mathbf{v}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i) \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{w}'_i \mathbf{M}_{\mathbf{f}_*^0} (-\bar{\mathbf{e}}_* \boldsymbol{\Lambda}_{*,y,i} + \mathbf{v}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i) \\
&\quad - \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{w}'_i (\mathbf{M}_{\mathbf{f}_*^0} - \mathbf{M}_{\hat{\mathbf{f}}^0}) (-\bar{\mathbf{e}}_* \boldsymbol{\Lambda}_{*,y,i} + \mathbf{v}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i) \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{w}'_i \mathbf{M}_{\mathbf{f}_*^0} (-\bar{\mathbf{e}}_* \boldsymbol{\Lambda}_{*,y,i} + \mathbf{v}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i) + O_p(N_0^{-1/2}) \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{w}'_i^{d'} (-\bar{\mathbf{e}}_*^d \boldsymbol{\Lambda}_{*,y,i} + \mathbf{v}_i^d \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i^d) + O_p(N_0^{-1/2}),
\end{aligned} \tag{A.95}$$

and so

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \boldsymbol{\Sigma}_{\mathbf{w}^d}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{w}_i^{d'} (-\bar{\mathbf{e}}_*^d \boldsymbol{\Lambda}_{*,y,i} + \mathbf{v}_i^d \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i^d) + o_p(1). \tag{A.96}$$

Together with the fact that $\hat{\mathbf{f}}_t^0 = \mathbf{f}_{*,t}^0 + O_p(N_0^{-1/2})$ by the arguments of Proof of Theorem 1,

the above result implies

$$\begin{aligned}
\widehat{\Delta}_{i,t} &= \Delta_{i,t} - (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \mathbf{w}_{i,t} - \boldsymbol{\Lambda}'_{*,y,i} \bar{\mathbf{e}}_{*,t} + \boldsymbol{\beta}'_i \mathbf{v}_{i,t} + \varepsilon_{i,t} \\
&\quad - [-\mathbf{w}_i(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \bar{\mathbf{e}}_* \boldsymbol{\Lambda}_{*,y,i} + \mathbf{v}_i \boldsymbol{\beta}_i + \varepsilon_i] \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0 \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 \\
&= \Delta_{i,t} - (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \mathbf{w}_{i,t} - \boldsymbol{\Lambda}'_{*,y,i} \bar{\mathbf{e}}_{*,t} + \boldsymbol{\beta}'_i \mathbf{v}_{i,t} + \varepsilon_{i,t} \\
&\quad - [-\mathbf{w}_i(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \bar{\mathbf{e}}_* \boldsymbol{\Lambda}_{*,y,i} + \mathbf{v}_i \boldsymbol{\beta}_i + \varepsilon_i] \mathbf{f}_*^0 (\mathbf{f}_*^0 \mathbf{f}_*^0)^{-1} \mathbf{f}_{*,t}^0 + O_p(N_0^{-1/2}) \\
&= \Delta_{i,t} - (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \mathbf{w}_{i,t}^d - \boldsymbol{\Lambda}'_{*,y,i} \bar{\mathbf{e}}_{*,t}^d + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d + O_p(N_0^{-1/2}),
\end{aligned} \tag{A.97}$$

giving

$$\begin{aligned}
&\sqrt{N_g} (\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} [v_{i,t} - (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \mathbf{w}_{i,t}^d - \boldsymbol{\Lambda}'_{*,y,i} \bar{\mathbf{e}}_{*,t}^d + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d] + O_p(N_0^{-1/2}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) - \sqrt{\frac{N_g}{N}} \sqrt{N} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \mathbf{w}_{i,t}^d \\
&\quad - \sqrt{\frac{N_g}{N_0} \frac{1}{N_g}} \sum_{i \in \mathcal{I}_g} \boldsymbol{\Lambda}'_{*,y,i} \sqrt{N_0} \bar{\mathbf{e}}_{*,t}^d + O_p(N_0^{-1/2}),
\end{aligned} \tag{A.98}$$

where the first and third terms are mean zero and asymptotically normal under the conditions of the main paper. The same is true for the second term, provided that \mathbf{w}_i is independent of \mathbf{v}_i and ε_i . Note in particular that if we assume as in Theorems 2 and 3 that $N_g/N_0 \rightarrow 0$, then we also have $N_g/N \rightarrow 0$, as $N > N_0$. The second and third terms are therefore negligible, and hence

$$\sqrt{N_g} (\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) = \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) + o_p(1). \tag{A.99}$$

We can therefore show that, up to the difference of the definition of the factors used in the defactoring, the presence of conditionally exogenous regressors has no effect on the asymptotic properties of $\sqrt{N_g} (\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t})$. The results in (a) and (b) are implied by this. \blacksquare

We end this section with a few remarks. We start with a discussion of how to relax the common slope condition. If the coefficient on $\mathbf{w}_{i,t}$ is not $\boldsymbol{\theta}$ but $\boldsymbol{\theta}_i$, we redefine

$$\widehat{y}_{i,t}(0) = \widehat{\boldsymbol{\theta}}'_i \mathbf{w}_{i,t} + \widehat{\mathbf{a}}'_i \widehat{\mathbf{f}}_t, \tag{A.100}$$

where $\widehat{\boldsymbol{\theta}}_i = (\mathbf{w}'_i \mathbf{M}_{\widehat{\mathbf{f}}} \mathbf{w}_i)^{-1} \mathbf{w}'_i \mathbf{M}_{\widehat{\mathbf{f}}} \mathbf{y}_i$. The effect of this change is easily appreciated given the results provided earlier. Note in particular how

$$\begin{aligned} (\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)' \mathbf{w}_{i,t}^d &= (-\bar{\mathbf{e}}_w^d \boldsymbol{\Lambda}_{*,y,i} + \mathbf{v}_i^d \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i^d)' \mathbf{w}_i^d (\mathbf{w}_i^{d'} \mathbf{w}_i^d)^{-1} \mathbf{w}_{i,t}^d \\ &= (\mathbf{v}_i^d \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i^d)' \mathbf{w}_i^d (\mathbf{w}_i^{d'} \mathbf{w}_i^d)^{-1} \mathbf{w}_{i,t}^d + \boldsymbol{\Lambda}'_{*,y,i} \bar{\mathbf{e}}_w^{d'} \mathbf{w}_i^d (\mathbf{w}_i^{d'} \mathbf{w}_i^d)^{-1} \mathbf{w}_{i,t}^d \\ &= (\mathbf{v}_i^d \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i^d)' \mathbf{w}_i^d (\mathbf{w}_i^{d'} \mathbf{w}_i^d)^{-1} \mathbf{w}_{i,t}^d + \text{tr} [\bar{\mathbf{e}}_*^{d'} \mathbf{w}_i^d (\mathbf{w}_i^{d'} \mathbf{w}_i^d)^{-1} \mathbf{w}_{i,t}^d \boldsymbol{\Lambda}'_{*,y,i}], \end{aligned} \quad (\text{A.101})$$

leading to the following limiting representation for $\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t})$:

$$\begin{aligned} &\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) \\ &= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \boldsymbol{\beta}'_i \mathbf{v}_{i,t}^d + \boldsymbol{\varepsilon}_{i,t}^d) - \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (\mathbf{v}_i^d \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i^d)' \mathbf{w}_i^d (\mathbf{w}_i^{d'} \mathbf{w}_i^d)^{-1} \mathbf{w}_{i,t}^d \\ &\quad + \sqrt{\frac{N_g}{N_0}} \text{tr} \left(\sqrt{N_0} \bar{\mathbf{e}}_*^{d'} \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \mathbf{w}_i^d (\mathbf{w}_i^{d'} \mathbf{w}_i^d)^{-1} \mathbf{w}_{i,t}^d \boldsymbol{\Lambda}'_{*,y,i} \right) \\ &\quad - \sqrt{\frac{N_g}{N_0} \frac{1}{N_g}} \sum_{i \in \mathcal{I}_g} \boldsymbol{\Lambda}'_{*,y,i} \sqrt{N_0} \bar{\mathbf{e}}_{*,t}^d + o_p(1), \end{aligned} \quad (\text{A.102})$$

where the third and fourth terms are negligible under $N_g/N_0 \rightarrow 0$, but not the second (and the first). The asymptotic distribution is still normal but the variance increases because of non-negligible second term.

The above modification of the counterfactual estimation procedure makes it possible to include regressors that do not admit to a common factor representation ($\mathbf{w}_{i,t}$); however, it does not remove the need for regressors that do admit to such a representation ($\mathbf{x}_{i,t}$), because we still require that $m+1 \geq r$. That is, as far as the number of factors that can possibly be accounted for by the procedure is concerned, it is only the latter type of regressors that matter.

Interestingly, the factor-in-regressors condition can be dropped completely, but then at the cost of other conditions that may or may not be more restrictive. As a starting point, we take the study of 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

$$\mathbf{f}_t = \mathbf{B}' \boldsymbol{\Psi}_t \quad (\text{A.103})$$

where $\Psi_t = \mathbb{E}(\mathbf{z}_{i,t})$ is constant in i and \mathbf{B} is an arbitrary $(m+1) \times r$ matrix of constants. Unlike the factors-in-regressors condition, (A.103) is not testable. However, if it holds, $\widehat{\mathbf{f}}_t = \bar{\mathbf{z}}_t$ can be used to estimate an arbitrary number of factors. In order to illustrate this last point, note that if (A.103) holds, letting $\mathbf{a}_i = \mathbf{B}\boldsymbol{\alpha}_i$, the interactive effects can be written as

$$\boldsymbol{\alpha}'_i \mathbf{f}_t = \mathbf{a}'_i \Psi_t = \mathbf{a}'_i \widehat{\mathbf{f}}_t + \mathbf{a}'_i (\Psi_t - \widehat{\mathbf{f}}_t), \quad (\text{A.104})$$

where $\Psi_t - \widehat{\mathbf{f}}_t$ is negligible, as $\bar{\mathbf{z}}_t \rightarrow_p \mathbb{E}(\mathbf{z}_{i,t})$ as $N \rightarrow \infty$ under standard regulatory conditions.

3 Bootstrap

Bootstrap methods are often superior to asymptotic theory when it comes to approximating the exact distributions of estimators. And they are particularly convenient in situations when conventional “plug-in” standard errors and confidence intervals are not available. This last observation has led to a surge in the use of the bootstrap in the treatment effects literature (see, for example, Callaway and Karami, 2023. Chan and Kwok, 2022, Gobillon and Magnac, 2016, and Xu, 2017). However, most studies do not establish the asymptotically validity of the particular bootstrap that they are using. In fact, the only exception known to is Callaway and Sant’Anna (2021), who propose a multiplier-type bootstrap procedure. The problem with this procedure is that it is not suitable in the presence of interactive effects.

In this section, we build on a recent strand of literature concerned with bootstrapping factor-augmented panel data regressions (see, for example, De Vos and Stauskas, 2024, and Westerlund et al., 2019) to develop an asymptotically valid bootstrap procedure for the proposed TECCE estimator. The algorithm that we consider is similar to the one employed by Westerlund et al. (2019), and proceeds as follows:

1. The bootstrap sample $\{\mathbf{z}_{i,1}^*, \dots, \mathbf{z}_{i,T}^*\}_{i \in \mathcal{I}_0}$ is generated by drawing data for entire cross-sections $\mathbf{z}_{i,1}^*, \dots, \mathbf{z}_{i,T}^*$ with replacement from the original, treated sample.
2. Compute

$$\widehat{\text{ATT}}_{g,t}^* = \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \widehat{\Delta}_{i,t}^* = \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} [y_{i,t}^* - \widehat{y}_{i,t}^*(0)], \quad (\text{A.105})$$

where $\widehat{y}_{i,t}^*(0) = \widehat{\mathbf{a}}_i^{*\prime} \widehat{\mathbf{f}}_t$ is $\widehat{y}_{i,t}(0)$ with $\widehat{\mathbf{f}}_t$ held fixed and $\mathbf{z}_{i,t}$ replaced by $\mathbf{z}_{i,t}^* = [y_{i,t}^*, \mathbf{x}_{i,t}^*]'$.

3. Repeat steps 1 and 2 B times.

Theorem A.2 below shows that the asymptotic distribution of $\sqrt{N_g}(\widehat{\text{ATT}}_{g,t}^* - \widehat{\text{ATT}}_{g,t})$ is the same as that of $\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t})$, which means that the asymptotic distribution can be successfully replicated by the bootstrap.

Theorem A.2. *Under Assumptions 1–10, as $N_g, N_0 \rightarrow \infty$ with $N_g/N_0 \rightarrow 0$,*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}^*[\sqrt{N_g}(\widehat{\text{ATT}}_{g,t}^* - \widehat{\text{ATT}}_{g,t}) \leq x | \mathcal{C}] - \mathbb{P}[\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) \leq x | \mathcal{C}] \right| \rightarrow_p 0,$$

where the inequalities are interpreted element-wise, \mathbb{P}^* signifies the probability measure induced by the bootstrap sampling conditional on the original sample, and \mathcal{C} is the same sigma-field as in Proof of Theorem 1.

Proof: In this proof it is convenient to assume without loss of generality that there is just one treated group and also that the cross-sectional units are ordered such that units $\mathcal{I}_0 = \{1, \dots, N_0\}$ are untreated, while units $\mathcal{I}_1 = \{N_0 + 1, \dots, N\}$ are treated, so that $N_1 = N - N_0$. The bootstrap sampling can be stated in the following way:

$$\mathbf{z}_{i,t}^* = \mathbf{z}_{i^*,t}, \tag{A.106}$$

where i^* is an index drawn with replacement from \mathcal{I}_1 . The bootstrap permutations in (A.106) are modeled using the $1 \times N_0$ selection vector $\mathbf{w}_i = [0, \dots, 0, 1, 0, \dots, 0]$, where the one is sitting in position i^* . This vector is drawn from a multinomial distribution with one trial and N_1 events with probability $1/N_1$. Define the $N_1 \times N_1$ matrix $\mathbf{w} = [\mathbf{w}'_{N_0+1}, \dots, \mathbf{w}'_N]'$ and the $N_1 \times 1$ vector $\mathbf{1}_{N_1} = [1, \dots, 1]'$. By construction, the $1 \times N_1$ vector $\mathbf{1}'_{N_1} \mathbf{w} = \sum_{i=N_0+1}^N \mathbf{w}_i = \mathbf{s} = [s_{N_0+1}, \dots, s_N]$ gives the number of times each cross-section unit in the original sample appears in the bootstrap sample; that is, s_i counts the number of times unit i appears in the bootstrap sample with $\sum_{i=N_0+1}^N s_i = N_1$. It follows that s_i is multinomially distributed with N_1 trials and N_1 events with probability $1/N_1$ (see, for example, De Vos and Stauskas, 2024, Supplement

A). Let $\mathbf{z}_t = [\mathbf{z}_{N_0+1,t}, \dots, \mathbf{z}_{N,t}]'$, a $N_1 \times m$ vector. In this notation,

$$\mathbf{z}_t^* = \begin{bmatrix} \mathbf{z}_{N_0+1,t}^{*'} \\ \vdots \\ \mathbf{z}_{N,t}^{*'} \end{bmatrix} = \mathbf{w}\mathbf{z}_t. \quad (\text{A.107})$$

Let $O_{p^*}(\cdot)$ be $O_p(\cdot)$ under the bootstrap-induced probability measure \mathbb{P}^* . By Lemma 3 of Cheng and Huang (2010), and the independence of s_i and $\mathbf{z}_{i,t}$, we have that $O_{p^*}(\cdot)$ in probability is equivalent to $O_p(\cdot)$ unconditionally. In what follows, we will therefore treat both weights and sample as random in the derivations. All orders hold in probability.

Analogously to Proof of Theorem 1, with $y_{i,t}^* = \eta_{i,t}^* + \beta_i^{*'} \mathbf{x}_{i,t}^* + \alpha_i^{*'} \mathbf{f}_t + \varepsilon_{i,t}^*$, $\hat{y}_{i,t}^*(0) = \hat{\mathbf{a}}_i^{*'} \hat{\mathbf{f}}_t$, $\hat{\mathbf{a}}_i^{*0} = (\bar{\mathbf{H}}\mathbf{D}_{N_0})^{-1} \hat{\mathbf{a}}_i^*$ and $\alpha_i^{*0} = (\bar{\mathbf{H}}\mathbf{D}_{N_0})^{-1} \mathbf{a}_i^* = [\alpha_i^{*'}, \mathbf{0}'_{(m+1-r) \times 1}]'$,

$$\begin{aligned} \hat{\Delta}_{i,t}^* &= y_{i,t}^* - \hat{y}_{i,t}^*(0) \\ &= \eta_{i,t}^* + \beta_i^{*'} \mathbf{x}_{i,t}^* - (\hat{\mathbf{a}}_i^{*'} \hat{\mathbf{f}}_t - \alpha_i^{*'} \mathbf{f}_t) + \varepsilon_{i,t}^* \\ &= \eta_{i,t}^* + \beta_i^{*'} \mathbf{x}_{i,t}^* - \alpha_i^{*0'} (\hat{\mathbf{f}}_t^0 - \mathbf{f}_t^0) - (\hat{\mathbf{a}}_i^{*0} - \alpha_i^{*0})' \hat{\mathbf{f}}_t^0 + \varepsilon_{i,t}^*. \end{aligned} \quad (\text{A.108})$$

In step 2 of the counterfactual estimation procedure, we have $i \in \mathcal{I}_1$ and $t \leq T_0$. This means that in the bootstrap world, $\mathbf{y}_i^* = \mathbf{x}_i^* \beta_i^* + \hat{\mathbf{f}} \bar{\mathbf{H}}_r \alpha_i^* - \bar{\mathbf{e}}_r^0 \alpha_i^* + \varepsilon_i^*$, which we can use to show that

$$\hat{\mathbf{a}}_i^* = (\hat{\mathbf{f}}' \hat{\mathbf{f}})^{-1} \hat{\mathbf{f}}' \mathbf{y}_i^* = \mathbf{a}_i^* + (\hat{\mathbf{f}}' \hat{\mathbf{f}})^{-1} \hat{\mathbf{f}}' (\mathbf{x}_i^* \beta_i^* - \bar{\mathbf{e}}_r^0 \alpha_i^* + \varepsilon_i^*), \quad (\text{A.109})$$

and so

$$\hat{\mathbf{a}}_i^{*0} = (\bar{\mathbf{H}}\mathbf{D}_{N_0})^{-1} \hat{\mathbf{a}}_i^* = \alpha_i^{*0} + (\hat{\mathbf{f}}^0' \hat{\mathbf{f}}^0)^{-1} \hat{\mathbf{f}}^0' (\mathbf{x}_i^* \beta_i^* - \bar{\mathbf{e}}_r^0 \alpha_i^* + \varepsilon_i^*). \quad (\text{A.110})$$

Moreover, by steps analogous to those used in the proof of Theorem 1, we obtain $\mathbf{x}_i^* = \hat{\mathbf{f}}^0 \mathbf{D}_{N_0}^{-1} \bar{\mathbf{H}}^{-1} \bar{\mathbf{H}}_r \lambda_i^* - \bar{\mathbf{e}}_r^0 \lambda_i^* + \mathbf{v}_i^*$ for $t \leq T_0$, and $\mathbf{x}_{i,t}^* = \tau_{i,t}^* + \lambda_i^{*'} \bar{\mathbf{H}}' \bar{\mathbf{H}}^{-1} \mathbf{D}_{N_0}^{-1} \hat{\mathbf{f}}_t^0 - \lambda_i^{*'} \bar{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t}^*$ for $t > T_0$. By inserting the above expressions into the one given earlier for $\hat{\Delta}_{i,t}^*$ and then using

$\widehat{\mathbf{f}}^0 = \mathbf{f}^+ + O_p(N_0^{-1/2})$ as in Proof of Theorem 1, we get

$$\begin{aligned}
\widehat{\Delta}_{i,t}^* &= \eta_{i,t}^* + \boldsymbol{\beta}_i^{*\prime} \mathbf{x}_{i,t}^* - \boldsymbol{\alpha}_i^{*\prime} \bar{\mathbf{e}}_{r,t}^0 + \varepsilon_{i,t}^* - (\mathbf{x}_i^* \boldsymbol{\beta}_i^* - \bar{\mathbf{e}}_r^0 \boldsymbol{\alpha}_i^* + \varepsilon_i^*)' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0 \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 \\
&= \eta_{i,t}^* + \boldsymbol{\beta}_i^{*\prime} (\boldsymbol{\tau}_{i,t}^* - \lambda_i^{*\prime} \bar{\mathbf{e}}_{r,t}^0 + \mathbf{v}_{i,t}^*) - \boldsymbol{\alpha}_i^{*\prime} \bar{\mathbf{e}}_{r,t}^0 + \varepsilon_{i,t}^* \\
&\quad - [(-\bar{\mathbf{e}}_r^0 \boldsymbol{\lambda}_i^* + \mathbf{v}_i^*) \boldsymbol{\beta}_i^* - \bar{\mathbf{e}}_r^0 \boldsymbol{\alpha}_i^* + \varepsilon_i^*]' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0 \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 \\
&= \Delta_{i,t}^* - \boldsymbol{\Lambda}_{y,i}^{*\prime} \bar{\mathbf{e}}_{r,t}^0 + \boldsymbol{\beta}_i^{*\prime} \mathbf{v}_{i,t}^* + \varepsilon_{i,t}^* - (-\bar{\mathbf{e}}_r^0 \boldsymbol{\Lambda}_{y,i}^* + \mathbf{v}_i^* \boldsymbol{\beta}_i^* + \varepsilon_i^*)' \widehat{\mathbf{f}}^0 (\widehat{\mathbf{f}}^0 \widehat{\mathbf{f}}^0)^{-1} \widehat{\mathbf{f}}_t^0 \\
&= \Delta_{i,t}^* - \boldsymbol{\Lambda}_{y,i}^{*\prime} \bar{\mathbf{e}}_{r,t}^0 + \boldsymbol{\beta}_i^{*\prime} \mathbf{v}_{i,t}^* + \varepsilon_{i,t}^* \\
&\quad - (-\bar{\mathbf{e}}_r^0 \boldsymbol{\Lambda}_{y,i}^* + \mathbf{v}_i^* \boldsymbol{\beta}_i^* + \varepsilon_i^*)' \mathbf{f}^+ (\mathbf{f}^{+\prime} \mathbf{f}^+)^{-1} \mathbf{f}_t^+ + O_{p^*}(N_0^{-1/2}) \\
&= \Delta_{i,t}^* - \boldsymbol{\Lambda}_{y,i}^{*\prime} \bar{\mathbf{e}}_{r,t}^{0d} + \boldsymbol{\beta}_i^{*\prime} \mathbf{v}_{i,t}^{*d} + \varepsilon_{i,t}^{*d} + O_{p^*}(N_0^{-1/2}),
\end{aligned} \tag{A.111}$$

which in view of $\Delta_{i,t}^* = \text{ATT}_{g,t} + v_{i,t}^*$ for $i \in \mathcal{I}_g$ gives

$$\begin{aligned}
&\sqrt{N_g} (\widehat{\text{ATT}}_{g,t}^* - \text{ATT}_{g,t}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (\widehat{\Delta}_{i,t}^* - \Delta_{i,t}^* + v_{i,t}^*) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t}^* - \boldsymbol{\Lambda}_{y,i}^{*\prime} \bar{\mathbf{e}}_{r,t}^{0d} + \boldsymbol{\beta}_i^{*\prime} \mathbf{v}_{i,t}^{*d} + \varepsilon_{i,t}^{*d}) + O_{p^*}(N_0^{-1/2}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t}^* + \boldsymbol{\beta}_i^{*\prime} \mathbf{v}_{i,t}^{*d} + \varepsilon_{i,t}^{*d}) - \sqrt{\frac{N_g}{N_0}} \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \boldsymbol{\Lambda}_{y,i}^{*\prime} \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} + O_{p^*}(N_0^{-1/2}) \\
&= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t}^* + \boldsymbol{\beta}_i^{*\prime} \mathbf{v}_{i,t}^{*d} + \varepsilon_{i,t}^{*d}) - \sqrt{\tau} \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \boldsymbol{\Lambda}_{y,i}^{*\prime} \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} + o_{p^*}(1),
\end{aligned} \tag{A.112}$$

where the last equality holds provided that $N_g/N_0 \rightarrow \tau < \infty$. Consider $N_g^{-1} \sum_{i \in \mathcal{I}_g} \boldsymbol{\Lambda}_{y,i}^{*\prime}$. Note how, in terms of the notation introduced in the beginning of this proof,

$$\sum_{i=N_0+1}^N \mathbf{z}_{i,t}^{*\prime} = \mathbf{1}'_{N_1} \mathbf{z}_t^* = \mathbf{1}'_{N_1} \mathbf{w} \mathbf{z}_t = \mathbf{s} \mathbf{z}_t = \sum_{i=N_0+1}^N s_i \mathbf{z}_{i,t}'. \tag{A.113}$$

This holds for all sums over bootstrapped quantities, including $\boldsymbol{\Lambda}_{y,i}^*$; that is, $\sum_{i \in \mathcal{I}_g} \boldsymbol{\Lambda}_{y,i}^* = \sum_{i \in \mathcal{I}_g} s_i \boldsymbol{\Lambda}_{y,i}$. Hence, since $\mathbb{E}(s_i) = 1$ by the properties of the multinomial distribution, we can show that

$$\frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \boldsymbol{\Lambda}_{y,i}^* = \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} s_i \boldsymbol{\Lambda}_{y,i} = \boldsymbol{\Lambda}_y + o_{p^*}(1), \tag{A.114}$$

where Λ_y is as in Proof of Theorem 1. It follows that

$$\begin{aligned} \sqrt{N_g}(\widehat{\text{ATT}}_{g,t}^* - \text{ATT}_{g,t}) &= \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} s_i(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) \\ &\quad - \sqrt{\tau} \Lambda'_y \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} + o_p(1). \end{aligned} \quad (\text{A.115})$$

Hence, since

$$\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) = \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) - \sqrt{\tau} \Lambda'_y \sqrt{N_0} \bar{\mathbf{e}}_{r,t}^{0d} + o_p(1) \quad (\text{A.116})$$

by Proof of Theorem 1, we have that

$$\sqrt{N_g}(\widehat{\text{ATT}}_{g,t}^* - \widehat{\text{ATT}}_{g,t}) = \frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (s_i - 1)(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) + o_p(1). \quad (\text{A.117})$$

The fact that s_i is independent of the sample data with $\mathbb{E}(s_i) = 1$ implies that the first term on the right-hand side is zero in expectation (conditionally on \mathcal{C}). Moreover, by the cross-sectional independence of $v_{i,t}$, $\mathbf{v}_{i,t}$ and $\varepsilon_{i,t}$,

$$\begin{aligned} \text{var} \left(\frac{1}{\sqrt{N_g}} \sum_{i \in \mathcal{I}_g} (s_i - 1)(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d) | \mathcal{C} \right) \\ = \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \sum_{j \in \mathcal{I}_g} \mathbb{E}[(s_i - 1)(s_j - 1)] \mathbb{E}[(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d)(v_{j,t} + \beta'_j \mathbf{v}_{j,t}^d + \varepsilon_{j,t}^d) | \mathcal{C}] \\ = \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \mathbb{E}[(s_i - 1)^2] \mathbb{E}[(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d)^2 | \mathcal{C}] \\ = \frac{1}{N_g} \sum_{i \in \mathcal{I}_g} \mathbb{E}[(v_{i,t} + \beta'_i \mathbf{v}_{i,t}^d + \varepsilon_{i,t}^d)^2 | \mathcal{C}] + O(N^{-1}) \end{aligned} \quad (\text{A.118})$$

where we have used the fact that $\mathbb{E}[(s_i - 1)^2] = 1 - N^{-1}$ by the properties of the multinomial distribution. Hence, by a central limit law for independent variables,

$$\sqrt{N_g}(\widehat{\text{ATT}}_{g,t}^* - \widehat{\text{ATT}}_{g,t}) \rightarrow_d MN(0, \text{var}(\widehat{\text{ATT}}_{g,t})) \quad (\text{A.119})$$

as $N_g, N_0 \rightarrow \infty$ with $\tau = 0$. From this result and Theorem 1 it follows that under the same conditions on N_g and N_0 that

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}^*[\sqrt{N_g}(\widehat{\text{ATT}}_{g,t}^* - \widehat{\text{ATT}}_{g,t}) \leq x | \mathcal{C}] - \mathbb{P}[\sqrt{N_g}(\widehat{\text{ATT}}_{g,t} - \text{ATT}_{g,t}) \leq x | \mathcal{C}] \right| \rightarrow_p 0,$$

which is what we wanted to show. ■

Theorem A.2 justifies the construction of bootstrap percentile confidence intervals (CIs) for the ATT with asymptotically correct coverage probabilities. Specifically, using $q(\alpha)$ to denote the $100 \cdot \alpha\%$ quantile of the distribution of $\widehat{\text{ATT}}_{g,t}^*$, the $100 \cdot (1 - \alpha)\%$ percentile bootstrap CI for $\text{ATT}_{g,t}$ is given by

$$\text{CI}(\alpha) = [q(\alpha/2), q(1 - \alpha/2)]. \quad (\text{A.120})$$

Theorem A.2 implies that $P[\text{ATT}_{g,t} \in \text{CI}(\alpha)] \rightarrow 1 - \alpha$ as $N_g, N_0 \rightarrow \infty$ with $\tau = 0$.

We can also bootstrap the statistics discussed in Section 4 of the main paper. For example, the bootstrap version of the $T_{g,t}$ -statistic considered for testing $H_0 : \text{ATT}_{g,t} = 0$ is given by

$$T_{g,t}^* := \frac{\sqrt{N_g}(\widehat{\text{ATT}}_{g,t}^* - \widehat{\text{ATT}}_{g,t})}{\sqrt{\widehat{\text{var}}(\widehat{\text{ATT}}_{g,t}^*)}}, \quad (\text{A.121})$$

where $\widehat{\text{var}}(\widehat{\text{ATT}}_{g,t}^*)$ is $\widehat{\text{var}}(\widehat{\text{ATT}}_{g,t})$ with $\widehat{\text{ATT}}_{g,t}$ and $\widehat{\Delta}_{i,t}$ replaced by $\widehat{\text{ATT}}_{g,t}^*$ and $\widehat{\Delta}_{i,t}^*$, respectively. The validity of $T_{g,t}^*$ is a direct consequence of Theorem A.2. This means that we may use the distribution of $T_{g,t}^*$ to compute asymptotically valid p -values for $T_{g,t}$ when testing H_0 . Let us therefore denote by $T_{g,t,b}^*$ the b -th bootstrap test statistic with $b = 1, \dots, B$. The bootstrapped p -value is given by $B^{-1} \sum_{b=1}^B 1(T_{g,t,b}^* > T_{g,t})$, where $1(A)$ is the indicator function for the event A , just as in the main paper.

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.
- CALLAWAY, B. AND S. KARAMI (2023): “Treatment effects in interactive fixed effects models with a small number of time periods,” *Journal of econometrics*, 233, 184–208.
- CALLAWAY, B. AND P. H. SANT’ANNA (2021): “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, S0304407620303948.
- 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.
- CHENG, G. AND J. Z. HUANG (2010): “Bootstrap Consistency for General Semiparametric M-estimation,” *The Annals of Statistics*, 38, 2884–2915.
- DE VOS, I. AND O. STAUSKAS (2024): “Cross-section bootstrap for CCE regressions,” *Journal of Econometrics*, 240, 105648.
- DE VOS, I. AND J. WESTERLUND (2019): “On CCE estimation of factor-augmented models when regressors are not linear in the factors,” *Economics letters*, 178, 5–7.
- GOBILLON, L. AND T. MAGNAC (2016): “Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls,” *Review of Economics and Statistics*, 98, 535–551.
- 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.

XU, Y. (2017): "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models," *Political Analysis*, 25, 57–76.