Online Appendix to: 'Dynamic Treatment Effect Estimation with Interactive Fixed Effects and Short Panels'

Nicholas Brown* and Kyle Butts†

July 7, 2022

A— Proofs

Proof of Theorem 2.1

Let $t \geq g$ for the given group g.

$$\mathbb{E}\big[y_{it} - P(f_t', F_{t<g})y_{i,t<g} \mid G_i = g\big] = \mathbb{E}\big[y_{it}(1) \mid G_i = g\big] - \mathbb{E}\big[P(f_t', F_{t<g})y_{i,t<g} \mid G_i = g\big]$$

We use the fact that

$$
\begin{aligned}
\mathbb{E}\big[P(f_t', F_{t<g})y_{i,t<g} \mid G_i = g\big] &= \mathbb{E}\big[f_t'(F_{t<g}'F_{t<g})^{-1}F_{t<g}'y_{i,t<g} \mid G_i = g\big] \\
&= \mathbb{E}\big[f_t'(F_{t<g}'F_{t<g})^{-1}F_{t<g}'\big[F_{t<g}\gamma_i + u_{i,t<g}\big] \mid G_i = g\big] \\
&= \mathbb{E}\big[f_t'\gamma_i + f_t'(F_{t<g}'F_{t<g})^{-1}F_{t<g}'u_{i,t<g} \mid G_i = g\big] \\
&= \mathbb{E}[y_{it}(\infty) \mid G_i = g]
\end{aligned}
$$

The second equality hold by Assumption 2 and the fact that $y_{i,t<g} = y_{i,t<g}(0)$. The final equality holds by Assumption 2.

□

*Queen's University, Economics Department (n.brown@queensu.ca)
†University of Colorado Boulder, Economics Department (kyle.butts@colorado.edu)

Proof of Theorem 2.2

There exists a $m \times p$ matrix $\boldsymbol{A}$ such that

$$\boldsymbol{F}(\boldsymbol{\theta})\boldsymbol{A} = \boldsymbol{F} \tag{A.1}$$

$\boldsymbol{A}$ defines the linear combinations of the columns of $\boldsymbol{F}(\boldsymbol{\theta})$ that span the columns of $\boldsymbol{F}$. Thus $\boldsymbol{f}(\boldsymbol{\theta})'_t \boldsymbol{A} = \boldsymbol{f}'_t$. We then have

$$\begin{aligned}
\boldsymbol{f}(\boldsymbol{\theta})'_t (\boldsymbol{F}(\boldsymbol{\theta})'_{t<g} \boldsymbol{F}(\boldsymbol{\theta})_{t<g})^{-1} \boldsymbol{F}(\boldsymbol{\theta})'_{t<g} \boldsymbol{F}_{t<g} \gamma_i &= \boldsymbol{f}(\boldsymbol{\theta})'_t (\boldsymbol{F}(\boldsymbol{\theta})'_{t<g} \boldsymbol{F}(\boldsymbol{\theta})_{t<g})^{-1} \boldsymbol{F}(\boldsymbol{\theta})'_{t<g} \boldsymbol{F}(\boldsymbol{\theta})_{t<g} \boldsymbol{A} \gamma_i \\
&= \boldsymbol{f}(\boldsymbol{\theta})'_t \boldsymbol{A} \gamma_i \\
&= \boldsymbol{f}'_t \gamma_i
\end{aligned}$$

If $m = p$ so that $\boldsymbol{F}$ also has full column rank, we can make the stronger statement that the imputation matrices of $\boldsymbol{F}$ and $\boldsymbol{F}(\boldsymbol{\theta})$ are equal:

$$\begin{aligned}
\boldsymbol{P}(\boldsymbol{F}_{t\geq g}, \boldsymbol{F}_{t<g}) &= \boldsymbol{F}_{t\geq g} (\boldsymbol{F}'_{t<g} \boldsymbol{F}_{t<g})^{-1} \boldsymbol{F}'_{t<g} \\
&= \boldsymbol{F}_{t\geq g} \boldsymbol{A} (\boldsymbol{A}' \boldsymbol{F}'_{t<g} \boldsymbol{F}_{t<g} \boldsymbol{A})^{-1} \boldsymbol{A}' \boldsymbol{F}'_{t<g} \\
&= \boldsymbol{F}(\boldsymbol{\theta})_{t\geq g} (\boldsymbol{F}(\boldsymbol{\theta})'_{t<g} \boldsymbol{F}(\boldsymbol{\theta})_{t<g})^{-1} \boldsymbol{F}(\boldsymbol{\theta})'_{t<g} \\
&= \boldsymbol{P}(\boldsymbol{F}(\boldsymbol{\theta})_{t\geq g}, \boldsymbol{F}(\boldsymbol{\theta})_{t<g})
\end{aligned}$$

where the second equality holds because $\boldsymbol{A}$ and $(\boldsymbol{F}'_{t<g} \boldsymbol{F}_{t<g})$ are full rank.

□

Proof of Lemma 2.1

We first derive the averages defined in Section 2.2 in terms of the potential outcome framework:

$$\overline{y}_{\infty,t} = \frac{1}{N_\infty}\sum_{i=1}^{N}D_{i\infty}y_{it} = \overline{\mu}_\infty + \lambda_t + \boldsymbol{f}_t\overline{\gamma}_\infty + \overline{u}_{t,\infty}$$

$$\overline{y}_{i,t\leq T_0} = \frac{1}{T_0}\sum_{t=1}^{T_0}y_{it} = \mu_i + \overline{\lambda}_{t<T_0} + \overline{\boldsymbol{f}}_{t<T_0}\gamma_i + \overline{u}_{i,t<T_0}$$

$$\overline{y}_{\infty,t<T_0} = \frac{1}{N_\infty T_0}\sum_{i=1}^{N}\sum_{t=1}^{T_0}D_{i\infty}y_{it} = \overline{\mu}_\infty + \overline{\lambda}_{t<T_0} + \overline{\boldsymbol{f}}_{t<T_0}\overline{\gamma}_\infty + \overline{u}_{\infty,t<T_0}$$

where $\overline{\mu}_\infty$ and $\overline{\gamma}_\infty$ are the averages of the never-treated individuals' heterogeneity and $\overline{\boldsymbol{f}}_{t<T_0}$ and $\overline{\lambda}_{t<T_0}$ are the averages of the time effects before anyone is treated. The error averages have the same interpretation as the outcome averages.

The definition of τ_{it} is the difference between treated and untreated potential outcomes for unit i at time t, so for any (i,t), $y_{it} = d_{it}y_{it}(1) + (1-d_{it})y_{it}(\infty) = d_{it}\tau_{it} + y_{it}(\infty)$. Then

$$\check{y}_{it} = d_{it}\tau_{it} + \boldsymbol{f}_t'\gamma_i - \overline{\boldsymbol{f}}_{t<T_0}'\gamma_i - \boldsymbol{f}_t'\overline{\gamma}_\infty + \overline{\boldsymbol{f}}_{t<T_0}\overline{\gamma}_\infty + u_{it} - \overline{u}_{t,\infty} - \overline{u}_{i,t<T_0} + \overline{u}_{\infty,t<T_0}$$

$$= d_{it}\tau_{it} + (\boldsymbol{f}_t - \overline{\boldsymbol{f}}_{t<T_0})'(\gamma_i - \overline{\gamma}_\infty) + u_{it} - \overline{u}_{t,\infty} - \overline{u}_{i,t<T_0} + \overline{u}_{\infty,t<T_0}$$

Taking expectation conditional on $G_i = g$ gives $\mathbb{E}\big[u_{it} - \overline{u}_{i,t<T_0} \mid G_i = g\big] = 0$ by Assumption 2 and $\mathbb{E}\big[\overline{u}_{\infty,t<T_0} - \overline{u}_{t,\infty} \mid G_i = g\big] = \mathbb{E}\big[\overline{u}_{\infty,t<T_0} - \overline{u}_{t,\infty}\big] = 0$ by random sampling and iterated expectations.

$\square$

Proof of Theorem 3.1

Asymptotic normality is a consequence of well-known large sample GMM theory. See, for example, Hansen (1982).

We only need to derive the asymptotic variances. Note that $\boldsymbol{g}_{i\infty}(\boldsymbol{\theta}) \otimes \boldsymbol{g}_{ig}(\boldsymbol{\theta}, \boldsymbol{\tau}_g) = \boldsymbol{0}$ (from the D_{ig} terms) and $\boldsymbol{g}_{ih}(\boldsymbol{\theta}, \boldsymbol{\tau}_h) \otimes \boldsymbol{g}_{ik}(\boldsymbol{\theta}, \boldsymbol{\tau}_k) = \boldsymbol{0}$ almost surely uniformly over the parameter space for all $g \in \mathcal{G}$ and $h \neq k$. The covariance matrix of these moment functions, which we denote as

$\boldsymbol{\Delta}$, is a block diagonal matrix.

$$
\boldsymbol{\Delta} =
\begin{pmatrix}
\mathbb{E}\left[g_{i\infty}(\boldsymbol{\theta})g_{i\infty}(\boldsymbol{\theta})'\right] & 0 & 0 & \dots & 0 \\
0 & \mathbb{E}\left[g_{ig_G}(\boldsymbol{\theta},\boldsymbol{\tau}_{g_G})g_{ig_G}(\boldsymbol{\theta},\boldsymbol{\tau}_{g_G})'\right] & 0 & \dots & 0 \\
\vdots & & & \ddots & \\
0 & 0 & 0 & \dots & \mathbb{E}\left[g_{ig_1}(\boldsymbol{\theta},\boldsymbol{\tau}_{g_1})g_{ig_1}(\boldsymbol{\theta},\boldsymbol{\tau})'\right]
\end{pmatrix}
$$

We write the individual blocks as $\boldsymbol{\Delta}_g$ for $g \in \mathcal{G} \cup \{\infty\}$. The gradient is also simple to compute because all of the moments are linear in the treatment effects. We define the overall gradient $\boldsymbol{D}$ and show it is a lower triangular matrix which we write in terms of its constituent blocks:

$$
\boldsymbol{D} =
\begin{pmatrix}
\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}g_{i\infty}(\boldsymbol{\theta})\right] & 0 & 0 & \dots & 0 \\
\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}g_{ig_G}(\boldsymbol{\theta},\boldsymbol{\tau}_{g_G})\right] & -\boldsymbol{I}_{T-g_G+1} & 0 & \dots & 0 \\
\vdots & & & \ddots & \\
\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}g_{ig_1}(\boldsymbol{\theta},\boldsymbol{\tau}_{g_1})\right] & 0 & 0 & \dots & -\boldsymbol{I}_{T-g_1+1}
\end{pmatrix}
$$

where we write the blocks in the first column as $\boldsymbol{D}_g$ for $g \in \mathcal{G} \cup \{\infty\}$. The diagonal is made up of negative identity matrices because $\mathbb{E}\left[\frac{D_{ig_h}}{\mathbb{P}(D_{ig_h}=1)}\right] = 1$.

Given we use the optimal weight matrix, the overall asymptotic variance is given by $(\boldsymbol{D}'\boldsymbol{\Delta}^{-1}\boldsymbol{D})^{-1}$. $\boldsymbol{\Delta}$ is a block diagonal matrix so its inverse is trivial to compute. First, we have

$$
\boldsymbol{\Delta}^{-1}\boldsymbol{D} =
\begin{pmatrix}
\boldsymbol{\Delta}_{\infty}^{-1}\boldsymbol{D}_{\infty} & 0 & \dots & 0 \\
\boldsymbol{\Delta}_{g_G}^{-1}\boldsymbol{D}_{g_G} & -\boldsymbol{\Delta}_{g_G}^{-1} & \dots & 0 \\
\vdots & & \ddots & \\
\boldsymbol{\Delta}_{g_1}^{-1}\boldsymbol{D}_{g_1} & 0 & \dots & -\boldsymbol{\Delta}_{g_1}^{-1}
\end{pmatrix}
$$

The transpose of the gradient matrix is

$$
D' = \begin{pmatrix}
D'_\infty & D'_{g_G} & \cdots & D'_{g_1} \\
0 & -I_{T-g_G+1} & \cdots & 0 \\
\vdots & & \ddots & \\
0 & 0 & \cdots & -I_{T-g_1+1}
\end{pmatrix}
$$

so that we get

$$
D'\Delta^{-1}D = \begin{pmatrix}
\sum_{g\in\mathcal{G}\cup\{\infty\}} D'_g\Delta_g^{-1}D_g & -D'_{g_G}\Delta_{g_G}^{-1} & \cdots & -D'_{g_1}\Delta_{g_G}^{-1} \\
-\Delta_{g_G}^{-1}D_{g_G} & \Delta_{g_G}^{-1} & \cdots & 0 \\
\vdots & & \ddots & \\
-\Delta_{g_1}^{-1}D_{g_1} & 0 & \cdots & \Delta_{g_1}^{-1}
\end{pmatrix}
$$

We write this matrix as

$$
\begin{pmatrix}
A & B \\
C & D
\end{pmatrix}
$$

where $A = \sum_{g\in\mathcal{G}\cup\{\infty\}} D'_g\Delta_g^{-1}D_g$ and $D = \mathrm{diag}\{\Delta_g^{-1}\}_{g\in\mathcal{G}}$. We then apply Exercise 5.16 of Abadir and Magnus (2005) to get the final inverse. The top left corner of the inverse is F^{-1} where

$$
(F)^{-1} = (A - BD^{-1}C)^{-1}
$$

$$
= \left(\sum_{g\in\mathcal{G}\cup\{\infty\}} D'_g\Delta_g^{-1}D_g - \left(\sum_{g\in\mathcal{G}} D'_g\Delta_g^{-1}D_g \right) \right)^{-1}
$$

$$
= (D'_\infty\Delta_\infty^{-1}D_\infty)^{-1}
$$

$$
= \mathrm{Avar}(\sqrt{N}(\widehat{\theta} - \theta))
$$

The rest of the first column of matrices takes the form

$$-D^{-1}CF^{-1} = \begin{pmatrix} D_{g_G} \\ \vdots \\ D_{g_1} \end{pmatrix} (D'_\infty \Delta_\infty^{-1} D_\infty)^{-1}$$

$$= \begin{pmatrix} D_{g_G}(D'_\infty \Delta_\infty^{-1} D_\infty)^{-1} \\ \vdots \\ D_{g_1}(D'_\infty \Delta_\infty^{-1} D_\infty)^{-1} \end{pmatrix}$$

and the rest of the first row is $-F^{-1}BD^{-1} = (-D^{-1}B'F^{-1})' = (-D^{-1}CF^{-1})'$.

Finally, the bottom-right block, which also gives the asymptotic covariance matrix of the ATT estimators, is

$$D^{-1} + D^{-1}CF^{-1}BD^{-1} = D^{-1} + \begin{pmatrix} D_{g_G}(D'_\infty \Delta_\infty^{-1} D_\infty)^{-1} D'_{g_G} & \cdots & D_{g_G}(D'_\infty \Delta_\infty^{-1} D_\infty)^{-1} D'_{g_1} \\ & \ddots & \\ D_{g_1}(D'_\infty \Delta_\infty^{-1} D_\infty)^{-1} D'_{g_G} & \cdots & D_{g_1}(D'_\infty \Delta_\infty^{-1} D_\infty)^{-1} D'_{g_1} \end{pmatrix}$$

The g'th diagonal elements of the resulting matrix is $\Delta_g + D_g(D'_\infty \Delta_\infty^{-1} D_\infty)^{-1} D'_g$.

$\square$

Proof of Theorem 3.2

We derive the limiting theory by multiplying $\widehat{\Delta}_g$ by $(N_g - 1)/N_g$ which produces the same limit as $N \to \infty$. We write

$$\frac{N_g - 1}{N_g} \widehat{\Delta}_g = \frac{1}{N_g} \sum_{i=1}^N D_{ig} \widehat{\Delta}_{ig} \widehat{\Delta}'_{ig} - \widehat{\tau}_g \widehat{\tau}'_g$$

We already know that $\widehat{\tau}_g \xrightarrow{p} \tau_g$ by Theorem 3.1. Note that

$$
\frac{1}{N_g} \sum_{i=1}^{N} D_{ig} \widehat{\Delta}_{ig} \widehat{\Delta}'_{ig} = \left(\frac{1}{N_g} \sum_{i=1}^{N} D_{ig} y_{i,t \geq g} y'_{i,t \geq g} \right) - \left(\frac{1}{N_g} \sum_{i=1}^{N} D_{ig} y_{i,t \geq g} y'_{i,t<g} \right) P(F_{t \geq g}(\widehat{\theta}), F_{t<g}(\widehat{\theta}))'
$$

$$
- P(F_{t \geq g}(\widehat{\theta}), F_{t<g}(\widehat{\theta})) \left(\frac{1}{N_g} \sum_{i=1}^{N} D_{ig} y_{i,t<g} y'_{i,t \geq g} \right)
$$

$$
- P(F_{t \geq g}(\widehat{\theta}), F_{t<g}(\widehat{\theta})) \left(\frac{1}{N_g} \sum_{i=1}^{N} D_{ig} y_{i,t<g} y'_{i,t \geq g} \right) P(F_{t \geq g}(\widehat{\theta}), F_{t<g}(\widehat{\theta}))'
$$

Given $P(F_{t \geq g}(\widehat{\theta}), F_{t<g}(\widehat{\theta}))$ is equal to its infeasible counterpart $P(F_{t \geq g}, F_{t<g})$ plus a $O_p(N^{-1/2})$ term, Assumption 1 and the weak law of large numbers imply

$$
\frac{1}{N_g} \sum_{i=1}^{N} D_{ig} \widehat{\Delta}_{ig} \widehat{\Delta}'_{ig} - \widehat{\tau}_g \widehat{\tau}'_g \xrightarrow{p} \mathbb{E}\left[g_{ig}(\theta, \tau_g) \mid G_i = g \right] = \Delta_g
$$

The inverse exists with probability approaching one by Assumption 5.

□

B — Inference of Aggregate Treatment Effects

As in Callaway and Sant'Anna (2021), we can form aggregates of our group-time average treatment effects. For example, event-study type coefficients would average over the τ_{gt} where $t - g = e$ for some relative event-time e with weights proportional to group membership. Consider a general aggregate estimand δ which we define as a weighted average of $ATT(g, t)$:

$$
\delta = \sum_{g \in \mathcal{G}} \sum_{t > T_0} w(g, t) \tau_{gt} \tag{B.2}
$$

where the weights $w(g, t)$ are non-negative and sum to one. Table 1 of Callaway and Sant'Anna (2021) and the surrounding discussion describes various treatment effect aggregates and discuss explicit forms for the weights.

Our plug-in estimate for δ is given by $\widehat{\delta} = \sum_{g \in \mathcal{G}} \sum_{t > T_0} \widehat{w}(g, t) \widehat{\tau}_{gt}$. Inference on this term follows directly from Corollary 2 in Callaway and Sant'Anna (2021) if we have the influence function for our τ_{gt} estimates. Rewriting our moment equations in an asymptotically linear

form, we have:

$$\sqrt{N}\left((\widehat{\theta}',\widehat{\tau}')'-(\theta',\tau')'\right)=-\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(D'\Delta^{-1}D)^{-1}D'\Delta^{-1}g_i(\theta,\tau)\right)+o_p(1). \tag{B.3}$$

This form comes from the fact that the weight matrix is positive definite with probability approaching one[1]. The first term on the right-hand side is the influence function and hence inference on aggregate quantities follows directly. This result allows for use of the multiplier bootstrap to estimate standard errors in a computationally efficient manner.

C — Inference in Two-Way Fixed Effect Model

We derive the asymptotic distribution of our imputation estimator based off of the two-way error model in equation (1). First, we note that this estimator can be written in terms of the imputation matrix from Section 2. In particular, let $\mathbf{1}_t$ be a $T \times 1$ vector of ones up the t'th spot, with all zeros after. Define $\overline{y}_\infty = (\overline{y}_{\infty,1},...,\overline{y}_{\infty,T})'$ be the full vector of never-treated cross-sectional averages. Then our imputation transformation can be written as

$$\tilde{y}_i = \left[I_T - P(\mathbf{1}_T, \mathbf{1}_{T_0})\right](y_i - \overline{y}_\infty) \tag{C.4}$$

where the tth component of the above T-vector is

$$d_{it}\tau_{it} + \tilde{u}_{it}, \tag{C.5}$$

with $\tilde{u}_{it}$ is defined as the same transformation as $\tilde{y}_{it}$.

The imputation step of our estimator is a just-identified system of equations. As such, we do not need to worry about weighting in implementation and inference comes from standard theory of M-estimators. In fact, we have the following closed-form solution for the estimator of

---

1. This is a well-known expansion for analyzing the asymptotic properties of GMM estimators. See Chapter 14 of Wooldridge (2010) for example.

a group-time average treatment effect:

$$\widehat{\tau}_{gt} = \frac{1}{N_g} \sum_i D_{ig} \tilde{y}_{it},$$ (C.6)

where $N_{gt} = \sum_i D_{ig}$ is the number of units in group g.

The following theorem characterizes estimation under the two-way error model:

Theorem C. 1. Assume untreated potential outcomes take the form of the two-way error model given in equation (1). Suppose Assumptions 1 and 3 hold, as well as Assumption 2 with $\gamma_i = 0$. Then for all (g, t) with $g > t$, $\widehat{\tau}_{gt}$ is conditionally unbiased for $\mathbb{E}\left[\tau_{it} \mid D_{ig} = 1\right]$, has the linear form

$$\sqrt{N_g}\left(\widehat{\tau}_{gt} - \tau_{gt}\right) = \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N} D_{ig}(\tau_{it} - \tau_{gt} + u_{it} - \overline{u}_{i,t<T_0} - \overline{u}_{\infty,t} + \overline{u}_{\infty,t<T_0})$$ (C.7)

and

$$\sqrt{N_1}(\widehat{\tau}_{gt} - \tau_{gt}) \xrightarrow{d} N(0, V_1 + V_0)$$ (C.8)

as $N \to \infty$, where V_1 and V_0 are given below and $\tau_{gt} = \mathbb{E}\left[y_{it}(g) - y_{it}(\infty) \mid D_{ig} = 1\right]$ is the group-time average treatment effect (on the treated). ∎

Theorem (C. 1) demonstrates the simplicity of our imputation procedure under the two-way error model. While the general factor structure requires more care, estimation and inference will yield a similar result.

Proof of Theorem C. 1

The transformed post-treatment observations are

$$\tilde{y}_{it} = \tau_{it} + u_{it} - \overline{u}_{\infty,t} - \overline{u}_{i,t<T_0} + \overline{u}_{\infty,t<T_0}$$ (C.9)

To show unbiasedness, take expectation conditional on $D_{ig} = 1$. This expected value is

$$\mathbb{E}\left[\tau_{it} + u_{it} - \overline{u}_{i,t<T_0} - \overline{u}_{\infty,t} + \overline{u}_{\infty,t<T_0} \mid D_{ig} = 1\right] = \mathbb{E}\left[\tau_{it} \mid D_{ig} = 1\right]$$ (C.10)

by Assumption 2 and 3.

For consistency, note that averaging over the sample with $D_{ig} = 1$, subtracting τ_{gt}, and multiplying $\sqrt{N_g}$ gives

$$\sqrt{N_g}\left(\widehat{\tau}_{gt} - \tau_{gt}\right) = \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N} D_{ig}(\tau_{it} - \tau_{gt} + u_{it} - \bar{u}_{i,t<T_0}) + \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N} D_{ig}(-\bar{u}_{\infty,t} + \bar{u}_{\infty,t<T_0})$$

(C.11)

which is two normalized sums of uncorrelated iid sequences that have mean zero (by iterated expectations) and finite fourth moments.

Rewriting the second term in terms of the original averages $\frac{1}{N_\infty} \sum_{i=1}^{N} -u_{i,t} + \bar{u}_{i,t<T_0}$ gives:

$$\sqrt{N_g}\left(\widehat{\tau}_{gt} - \tau_{gt}\right) = \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N} D_{ig}(\tau_{it} - \tau_{gt} + u_{it} - \bar{u}_{i,t<T_0}) + \sqrt{\frac{N_g}{N_\infty}}\left(\frac{1}{\sqrt{N_\infty}} \sum_{i=1}^{N} D_{i\infty}(-u_{i,t} + \bar{u}_{i,t<T_0})\right)$$

(C.12)

Since these terms are mean zero and uncorrelated, we find the variance of each term separately.

The first term has asymptotic variance

$$V_1 = \mathbb{E}\left[\left(\tau_{it} - \tau_{gt} + u_{it} - \bar{u}_{i,t<T_0}\right)\left(\tau_{it} - \tau_{gt} + u_{it} - \bar{u}_{i,t<T_0}\right)' \mid D_{ig} = 1\right]$$

(C.13)

and the second term has asymptotic variance

$$V_0 = \frac{\mathbb{P}(D_{ig} = 1)}{\mathbb{P}(D_{i\infty} = 1)} \mathbb{E}\left[\left(\bar{u}_{i,t<T_0} - u_{i,t}\right)\left(\bar{u}_{i,t<T_0} - u_{i,t}\right)' \mid D_{i\infty} = 1\right]$$

(C.14)

The result follows from the independence of the two sums.

D — Including Covariates

We now discuss the inclusion of covariates in the untreated potential outcome mean model. Allowing for covariates further weakens our parallel trends assumption by allowing selection to hold on unobserved heterogeneity as well as observed characteristics. Identifying the effects of covariates requires some kind of time and unit variation because we manually remove the level

fixed effects.

A common inclusion in the treatment effects literature is time-constant variables with time-varying slopes. Suppose x_i is $1 \times K$ vector of time-constant covariates. We could write the mean model of the untreated outcomes as

$$\mathbb{E}\left[y_{it}(\infty) \mid x_i, \mu_i, \gamma_i, D_i\right] = x_i \beta_t + \mu_i + \lambda_t + f_t' \gamma_i \tag{D.15}$$

which allows observable covariates to have trending partial effects; covariates with constant slopes are captured by the unit effect. After removing the additive fixed effects, $x_i \beta_t$ will take the same form as the residuals of factor structure. Estimating θ can be done jointly with the time-varying coefficients by applying the QLD transformation to the vector of $\tilde{y}_{it} - \tilde{x}_i \tilde{\beta}_t$. We cannot identify the underlying partial effects because of the time-demeaning, but we can include them for the sake of strengthening the parallel trends assumption.

Time-constant covariates (or time-varying covariates fixed at their pre-treatment value) are often employed because there is little worry that they are affected by treatment. However, we could also include time- and individual-varying covariates of the form x_{it} that are allowed to have identifiable constant slopes if we assume their distribution is unaffected by treatment status. Let x_{it} be a $1 \times K$ vector of covariates that vary over i and t. We can jointly estimate a $K \times 1$ vector of parameters β along with θ using the moments

$$\mathbb{E}\left[H(\theta)'(\tilde{y}_i - \tilde{X}_i \beta) \otimes w_i \mid G_i = \infty\right] = 0 \tag{D.16}$$

where $\tilde{X}_i$ is the $T \times K$ matrix of stacked covariates after our double-demeaning procedure.

We could also allow slopes to vary across groups and estimate them via the group-specific pooled regression $D_{ig} y_{it}$ on $D_{ig} x_{it}$ with unit-specific slopes on $D_{ig} \tilde{f}(\widehat{\theta})_t$ for $t = 1, ..., g-1$. Then we include the covariates and their respective slopes into the moment conditions

$$\mathbb{E}\left[(\tilde{y}_{i,t \geq g} - \tilde{X}_{i,t \geq g} \beta_g) - P(\tilde{F}_{t \geq g}, \tilde{F}_{t<g})(\tilde{y}_{i,t<g} - \tilde{X}_{i,t<g} \beta_g) - \tau_g \mid G_i = g\right] = 0 \tag{D.17}$$

We note that the above expression requires treatment to not affect the evolution of the covariates, a strong assumption in practice. Chan and Kwok (2022) make a similar assumption for their

principal components difference-in-differences estimator. We study this assumption in the context of the common correlated effects model in Brown et al. (2023).

E — Testing Mean Equality of Factor Loadings

We develop this test in the context of the QLD estimation of Ahn et al. (2013). Specifically, we need $\mathbb{E}[\boldsymbol{\gamma}_i] = \mathbb{E}[\boldsymbol{\gamma}_i \mid G_i = g]$ for all $g \in \mathscr{G}$. Our imputation approach allows us to identify these terms up to a rotation. To see how, let $\boldsymbol{A}^*$ be the rotation that imposes the Ahn et al. (2013) normalization. Then

$$
\begin{aligned}
\boldsymbol{P}(\boldsymbol{I}_p, \boldsymbol{F}(\boldsymbol{\theta})_{t<g})\mathbb{E}\big[\boldsymbol{y}_{i,t<g} \mid G_i = g\big] &= \Big(\boldsymbol{F}(\boldsymbol{\theta})'_{t<g}\boldsymbol{F}(\boldsymbol{\theta})_{t<g}\Big)^{-1}\boldsymbol{F}(\boldsymbol{\theta})'_{t<g}\boldsymbol{F}_{t<g}\mathbb{E}[\boldsymbol{\gamma}_i \mid G_i = g] \\
&= \Big(\boldsymbol{F}(\boldsymbol{\theta})'_{t<g}\boldsymbol{F}(\boldsymbol{\theta})_{t<g}\Big)^{-1}\boldsymbol{F}(\boldsymbol{\theta})'_{t<g}\boldsymbol{F}(\boldsymbol{\theta})_{t<g}(\boldsymbol{A}^*)^{-1}\mathbb{E}[\boldsymbol{\gamma}_i \mid G_i = g] \\
&= (\boldsymbol{A}^*)^{-1}\mathbb{E}[\boldsymbol{\gamma}_i \mid G_i = g]
\end{aligned}
$$

where $\boldsymbol{F}(\boldsymbol{\theta}) = \boldsymbol{F}\boldsymbol{A}^*$.

It is irrelevant that the means of the factor loadings are only known up to a nonsingular transformation, because $\boldsymbol{A}^*$ is the same for each $g \in \mathscr{G}$ by virtue of the common factors. We note that

$$
\mathbb{E}[\boldsymbol{\gamma}_i \mid G_i = g] - \mathbb{E}[\boldsymbol{\gamma}_i] = \boldsymbol{0} \iff (\boldsymbol{A}^*)^{-1}(\mathbb{E}[\boldsymbol{\gamma}_i \mid G_i = g] - \mathbb{E}[\boldsymbol{\gamma}_i]) = \boldsymbol{0} \tag{E.18}
$$

The results above show how we can identify $(\boldsymbol{A}^*)^{-1}\mathbb{E}[\boldsymbol{\gamma}_i \mid G_i = g]$ by imputing the pre-treatment observations onto an identify matrix.

Collect the moments

$$\mathbb{E}\left[\frac{D_{i\infty}}{\mathbb{P}(D_{i\infty}=1)}H(\boldsymbol{\theta})\tilde{\boldsymbol{y}}_i \otimes \boldsymbol{w}_i\right]=0$$

$$\mathbb{E}\left[\frac{D_{i\infty}}{\mathbb{P}(D_{i\infty}=1)}\left(P(\boldsymbol{I}_p,\boldsymbol{F}(\boldsymbol{\theta}))\boldsymbol{y}_i-\boldsymbol{\gamma}^*\right)\right]=0$$

$$\mathbb{E}\left[\frac{D_{ig_G}}{\mathbb{P}(D_{ig_G}=1)}\left(P(\boldsymbol{I}_p,\boldsymbol{F}(\boldsymbol{\theta})_{t<g_G})\boldsymbol{y}_{i,t<g_G}-\boldsymbol{\gamma}^*_{g_G}\right)\right]=0$$

$$\vdots$$

$$\mathbb{E}\left[\frac{D_{ig_1}}{\mathbb{P}(D_{ig_1}=1)}\left(P(\boldsymbol{I}_p,\boldsymbol{F}(\boldsymbol{\theta})_{t<g_1})\boldsymbol{y}_{i,t<g_1}-\boldsymbol{\gamma}^*_{g_G}\right)\right]=0$$

The parameters $(\boldsymbol{\gamma}^*,\boldsymbol{\gamma}^*_{g_G},...,\boldsymbol{\gamma}^*_{g_1})$ represent the rotated means of the factor loadings. $\boldsymbol{\gamma}$ is the unconditional mean $(\boldsymbol{A}^*)^{-1}\mathbb{E}[\boldsymbol{\gamma}_i]$ and $\boldsymbol{\gamma}_g$ is the conditional mean $(\boldsymbol{A}^*)^{-1}\mathbb{E}[\boldsymbol{\gamma}_i \mid G_i = g]$ for $g \in \mathscr{G}$. We include estimation of the factors for convenience, so that one does not need to directly calculate the effect of first-stage estimation on the asymptotic variances of conditional means.

Joint GMM estimation of the above parameters, including $\boldsymbol{\theta}$, then allows one to test combinations of the rotated means. Specifically, we have the following result: **Nick:** Need to change the theorem reference/label

Theorem E. 1. If $\mathbb{E}[\boldsymbol{\gamma}_i \mid G_i = g] = \mathbb{E}[\boldsymbol{\gamma}_i]$ for all $g \in \mathscr{G}$, then

$$\boldsymbol{\gamma}^* = \boldsymbol{\gamma}^*_{g_G} = ... = \boldsymbol{\gamma}^*_{g_1} \tag{E.19}$$

∎

References

Abadir, Karim M., and Jan R. Magnus. 2005. *Matrix Algebra*. Volume 1. Cambridge University Press, . 10.1017/cbo9780511810800.

Ahn, Seung C, Young H Lee, and Peter Schmidt. 2013. "Panel data models with multiple time-varying individual effects." *Journal of econometrics* 174 (1): 1–14. 10.1016/j.jeconom.2012.12.002.

Brown, Nicholas, Kyle Butts, and Joakim Westerlund. 2023. "Difference-in-Differences via Common Correlated Effects." *arXiv preprint arXiv:2301.11358*, https://arxiv.org/abs/2301.11358.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics* 225 (2): 200–230. 10.1016/j.jeconom.2020.12.001.

Chan, Marc K, and Simon S Kwok. 2022. "The PCDID approach: difference-in-differences when trends are potentially unparallel and stochastic." *Journal of Business & Economic Statistics* 40 (3): 1216–1233. 10.1080/07350015.2021.1914636.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 1029–1054. 10.2307/1912775.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. MIT press.