Online Appendix to: 'A Unified Framework for Dynamic Treatment Effect Estimation in Interactive Fixed Effect Models'

Nicholas Brown* and Kyle Butts†

November 30, 2022

A — 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 $ATT(g, t)$:

$$\delta = \sum_{g \in \mathcal{G}} \sum_{t > T_0} w(g, t) \tau_{gt} \tag{OA.A.1}$$

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 $\hat{\delta} = \sum_{g \in \mathcal{G}} \sum_{t > T_0} \hat{w}(g, t) \hat{\tau}_{gt}$. Inference on this term follows directly from Corollary 2 in Callaway and Sant'Anna (2021) if we have

*Queen's University, Economics Department (n.brown@queensu.ca)

†University of Colorado Boulder, Economics Department (kyle.butts@colorado.edu)

the influence function for our τ_{gt} estimates. Rewriting our moment equations in an asymptotically linear form, we have:

$$\sqrt{N}\Big((\widehat{\boldsymbol{\theta}}', \widehat{\boldsymbol{\tau}}')' - (\boldsymbol{\theta}', \boldsymbol{\tau}')'\Big) = -\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\boldsymbol{D}'\boldsymbol{\Delta}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{\Delta}^{-1}\boldsymbol{g}_i(\boldsymbol{\theta}, \boldsymbol{\tau})\right) + o_p(1). \quad \text{(OA.A.2)}$$

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.

B — 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 **??**. In particular, let $\mathbf{1}_t$ be a $T \times 1$ vector of ones up the t'th spot, with all zeros after. Define $\overline{\boldsymbol{y}}_\infty = (\overline{y}_{\infty,1}, \dots, \overline{y}_{\infty,T})'$ be the full vector of never-treated cross-sectional averages. Then our imputation transformation can be written as

$$\tilde{\boldsymbol{y}}_i = \left[\boldsymbol{I}_T - \boldsymbol{P}(\mathbf{1}_T, \mathbf{1}_{T_0})\right](\boldsymbol{y}_i - \overline{\boldsymbol{y}}_\infty) \quad \text{(OA.B.3)}$$

where the t^{th} component of the above T-vector is

$$d_{it}\tau_{it} + \tilde{u}_{it}, \quad \text{(OA.B.4)}$$

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

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

the estimator of a group-time average treatment effect:

$$\widehat{\tau}_{gt} = \frac{1}{N_g}\sum_i D_{ig}\tilde{y}_{it}, \tag{OA.B.5}$$

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 OA.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 $\boldsymbol{\gamma}_i = 0$. Then for all (g,t) with $g > t$, $\widehat{\boldsymbol{\tau}}_{gt}$ is conditionally unbiased for $\mathbb{E}\left[\tau_{it} \mid D_{ig} = 1\right]$, has the linear form

$$\sqrt{N_g}\left(\widehat{\tau}_{gt} - \tau_{gt}\right) = \frac{1}{\sqrt{N_g}}\sum_{i=1}^{N} D_{ig}(\tau_{it} - \tau_{gt} + u_{it} - \overline{u}_{i,t<T_0} - \overline{u}_{\infty,t} + \overline{u}_{\infty,t<T_0}) \tag{OA.B.6}$$

and

$$\sqrt{N_1}(\widehat{\tau}_{gt} - \tau_{gt}) \xrightarrow{d} N(0, V_1 + V_0) \tag{OA.B.7}$$

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 (OA.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 OA.1

The transformed post-treatment observations are

$$\tilde{y}_{it} = \tau_{it} + u_{it} - \overline{u}_{\infty,t} - \overline{u}_{i,t<T_0} + \overline{u}_{\infty,t<T_0} \tag{OA.B.8}$$

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

$$\mathbb{E}\left[\tau_{it} + u_{it} - \bar{u}_{i,t<T_0} - \bar{u}_{\infty,t} + \bar{u}_{\infty,t<T_0} \mid D_{ig} = 1\right] = \mathbb{E}\left[\tau_{it} \mid D_{ig} = 1\right] \tag{OA.B.9}$$

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

$$\sqrt{N_g}\left(\hat{\tau}_{gt} - \tau_{gt}\right) = \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N} D_{ig}(\tau_{it} - \tau_{gt} + u_{it} - \bar{u}_{i,t<T_0}) + \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N} D_{ig}(-\bar{u}_{\infty,t} + \bar{u}_{\infty,t<T_0}) \tag{OA.B.10}$$

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:

$$\sqrt{N_g}\left(\hat{\tau}_{gt} - \tau_{gt}\right) = \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N} D_{ig}(\tau_{it} - \tau_{gt} + u_{it} - \bar{u}_{i,t<T_0}) + \sqrt{\frac{N_g}{N_\infty}}\left(\frac{1}{\sqrt{N_\infty}} \sum_{i=1}^{N} D_{i\infty}(-u_{i,t} + \bar{u}_{i,t<T_0})\right) \tag{OA.B.11}$$

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

The first term has asymptotic variance

$$V_1 = \mathbb{E}\left[\left(\tau_{it} - \tau_{gt} + u_{it} - \bar{u}_{i,t<T_0}\right)\left(\tau_{it} - \tau_{gt} + u_{it} - \bar{u}_{i,t<T_0}\right)' \mid D_{ig} = 1\right] \tag{OA.B.12}$$

and the second term has asymptotic variance

$$V_0 = \frac{\mathbb{P}(D_{ig} = 1)}{\mathbb{P}(D_{i\infty} = 1)} \mathbb{E}\left[\left(\bar{u}_{i,t<T_0} - u_{i,t}\right)\left(\bar{u}_{i,t<T_0} - u_{i,t}\right)' \mid D_{i\infty} = 1\right] \tag{OA.B.13}$$

The result follows from the independence of the two sums.

C — Testing Equality of Factors

An important assumption underlying our approach is that the factors, which affect both the pre- and post-treatment outcomes, are equal between the treated and untreated groups. This assumption may not hold if, for example, the control and treatment groups are geographically or sociologically separated. We, therefore, derive tests for equivalence of the factors.

We can only compare the pre-treatment factors because those are the ones the treated groups can identify. Testing each group sequentially may give misleading results, especially when there are few units per group. Therefore, we combine all treated groups and only compare the first T_0 factor observations before any group is treated. We define $D_i = \sum_{g \in \mathcal{G}} D_{ig}$ which is one if the unit is ever treated.

We consider two estimators of the pre-sample factors, one using the untreated observations and one using the pre-treated observations. The rank condition on $\boldsymbol{F}$ in Assumption 5(i) means we can hope to identify the pre-treatment factors with the pre-treatment treated observations. We apply the Ahn et al. (2013) normalization to the pre-treatment factors, and define $\boldsymbol{H}^*(\boldsymbol{\theta})' = \left[\boldsymbol{I}_{(T_0-p)} \boldsymbol{\Theta}^*\right]$ where $\boldsymbol{\Theta}^*$ is $(T_0 - p) \times p$ matrix of free parameters.

Given the appropriate identifying assumptions on the treated units, the two sets of moments are then

$$\mathbb{E}\left[\boldsymbol{g}_i^0(\boldsymbol{\theta}_0)\right] = \mathbb{E}\left[\frac{(1-D_i)}{\mathbb{P}(D_i = 0)} \boldsymbol{H}^*(\boldsymbol{\theta}_0)' \boldsymbol{y}_{i,t<T_0} \otimes \boldsymbol{w}_i\right] = \boldsymbol{0}_{T_0 \times 1}$$
$$\mathbb{E}\left[\boldsymbol{g}_i^1(\boldsymbol{\theta}_1)\right] = \mathbb{E}\left[\frac{D_i}{\mathbb{P}(D_i = 1)} \boldsymbol{H}^*(\boldsymbol{\theta}_1)' \boldsymbol{y}_{i,t<T_0} \otimes \boldsymbol{w}_i\right] = \boldsymbol{0}_{T_0 \times 1}$$

which are the unconditional versions of the moments based on both respective subsamples, and $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are the vectorizations of the $(T_0 - p) \times p$ unrestricted parameters associated with the ALS normalization applied to $\boldsymbol{F}_{\text{pre}}$. We write the empirical analogs as $\boldsymbol{g}^1(\boldsymbol{\theta}_j) = \frac{1}{N_1}\sum_{i=1}^{N} D_1 \boldsymbol{g}_i^1(\boldsymbol{\theta}_1)$ and $\boldsymbol{g}^0(\boldsymbol{\theta}_0) = \frac{1}{N_0}\sum_{i=1}^{N}(1-D_i)\boldsymbol{g}_i^0(\boldsymbol{\theta}_0)$ where N_0 and N_1 are the number of never-treated and treated individuals, respectively.

First, we must test whether the number of factors affecting both groups is the same. This can be achieved simply by estimating p separately using both subsamples. ALS provide tests for estimating p using their GMM estimator. Given that p is the same for both sets of moment conditions, we are interested in testing the null hypothesis $H_0 : \boldsymbol{\theta}_0 - \boldsymbol{\theta}_1 = \boldsymbol{0}$. This condition suffices for testing the equality of the pretreatment factors for the untreated and pre-treated groups, which we denote $\boldsymbol{F}_{0,\text{pre}}$ and $\boldsymbol{F}_{1,\text{pre}}$ respectively. This fact holds because

$$\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1 \iff \boldsymbol{F}(\boldsymbol{\theta}_0) = \boldsymbol{F}(\boldsymbol{\theta}_1) \iff \boldsymbol{F}(\boldsymbol{\theta}_0)\boldsymbol{A} = \boldsymbol{F}(\boldsymbol{\theta}_1)\boldsymbol{A} \iff \boldsymbol{F}_0 = \boldsymbol{F}_1 \tag{OA.C.14}$$

where the second equivalence holds because the rotation matrix $\boldsymbol{A}$ is nonsingular, just as in Section 4.

We define the variance matrices as

$$\boldsymbol{S}_0(\boldsymbol{\theta}_0) = Var(\boldsymbol{g}_i^0)$$
$$\boldsymbol{S}_1(\boldsymbol{\theta}_1) = Var(\boldsymbol{g}_i^1)$$

with consistent estimators $\widehat{\boldsymbol{S}}_0$ and $\widehat{\boldsymbol{S}}_1$. Let

$$\boldsymbol{J}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) = \frac{N_0}{N} g^0(\boldsymbol{\theta}_0)' \widehat{\boldsymbol{S}}_0^{-1} g^0(\boldsymbol{\theta}_0) + \frac{N_1}{N} g^1(\boldsymbol{\theta}_1)' \widehat{\boldsymbol{S}}_1^{-1} \boldsymbol{g}^1(\boldsymbol{\theta}_1) \tag{OA.C.15}$$

Finally, define $\widehat{\boldsymbol{\theta}}$ as the estimator of $\boldsymbol{\theta}$ which uses both sets of moment conditions, and let $\widehat{\boldsymbol{\theta}_0}$, $\widehat{\boldsymbol{\theta}}_1$ be the estimators using the respective subsamples and their respective moment conditions.

Theorem OA.1. Under Assumptions 1-5 and the null hypothesis,

$$N * \left(J(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}) - J(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{\theta}}_1)\right) \xrightarrow{d} \chi^2_{((T_0-p)p)} \tag{OA.C.16}$$

as $N \to \infty$. ■

This result is a direct application of Theorem 5.8 from Hall (2004). He requires the partial sums $\sqrt{N}\boldsymbol{g}^0(\boldsymbol{\theta}_0)$ and $\sqrt{N}\boldsymbol{g}^1(\boldsymbol{\theta}_1)$ be uncorrelated, which holds under random sampling. Further, we can replace N_0/N and N_1/N in equation (OA.C.15) with their asymptotic counterparts $\mathbb{P}(D_i = 0)$ and $\mathbb{P}(D_i = 1)$ because they are multiplied by $O_p(N^{-1})$ terms. $\mathbb{P}(D_i = 0)$ takes the place of π in Hall (2004).

References

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.

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

Hall, Alastair R. 2004. *Generalized Method of Moments*. OUP Oxford.

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