Differences-in-Differences Estimators of Intertemporal Treatment Effects with Periodically Missing Outcomes

Clément de Chaisemartin Diego Ciccia[*] Felix Knau Mélitine Malézieux

June 28, 2024

In what follows, we use the same notation and assumptions from Section 2 of de Chaisemartin and D'Haultfoeuille (2024). We have a panel of G groups whose treatment $D_{g,t}$ is observed for T periods. We want to estimate the effect of this treatment on an outcome $Y_{g,t}$ that is regularly observed once every $k > 1$ periods. The observation lag k is the same for all groups and it remains constant over time.

Periodically missing outcome. Let $\tau > 0$ be the first period where $Y_{g,t}$ is observed. $Y_{g,t}$ is not missing if and only if

$$(t - \tau) \bmod k = 0. \tag{1}$$

In other words, $Y_{g,t}$ is not missing if t is exactly $b \times k$ periods away from τ, for some nonnegative integer b. Notice that, by construction, $\tau \leq k$. If $\tau > k$, $Y_{g,\tau-k}$ is missing, since $Y_{g,\tau}$ is the first period where $Y_{g,t}$ is not missing. However, $Y_{g,\tau-k}$ cannot be missing since $[(\tau - k) - \tau] \bmod k = 0$. Hence, τ cannot be greater than k.

Collapsed data. Let $\widetilde{T} = \lfloor (T - \tau)/k \rfloor$ be the number of periods when $Y_{g,t}$ is not missing. From the original sample of $G \times T$ observations, we consider the $G \times \widetilde{T}$ subsample with only the periods where we observe $Y_{g,t}$. We will be referring to this subset as *collapsed* or $G \times \widetilde{T}$ data. Table 1 is an example of such data structure. Each row denotes a time period, while groups' treatment paths are represented in columns. The outcome variable is non-missing every 3 periods starting from period 3. As a result, the collapsed version of this toy dataset contains only 3 periods, i.e. $\tilde{t} = 3, 6, 9$.

We denote the function that maps $\tilde{t}$ in the collapsed data to the corresponding t in the original data as $p(\tilde{t}) \equiv \tau + (\tilde{t} - 1)k$ for all $\tilde{t} \in \{1, ..., \widetilde{T}\}$. Similarly, we denote its inverse as $p^{-1}(t) = 1 + (t - \tau)/k$ for all t such that Equation 1 is verified. Let

$$\widetilde{D}_{g,\tilde{t}} = D_{g,p(\tilde{t})} \qquad\qquad \widetilde{Y}_{g,\tilde{t}} = Y_{g,p(\tilde{t})} \tag{2}$$

be the treatment and outcome variables in the collapsed data, for $\tilde{t} = 1, ..., \widetilde{T}$. Finally, let

$$\widetilde{F}_g = \min_{\tilde{t} \in \{2,...,\widetilde{T}\}} \widetilde{D}_{g,\tilde{t}} \neq \widetilde{D}_{g,\tilde{t}-1} = \min_{p^{-1}(t): t \geq 2, (t-\tau) \bmod k = 0} D_{g,t} \neq D_{g,t-k} \tag{3}$$

[*]Corresponding author. Email contact: chaisemartin.packages@gmail.com

T	$\widetilde{T}$	$D_{1,t}$	$D_{2,t}$	$D_{3,t}$	$D_{4,t}$	$D_{5,t}$
1	.	0	0	0	0	0
2	.	0	1	0	0	0
3	1	0	0	0	0	0
4	.	1	0	0	0	0
5	.	0	0	0	0	0
6	2	0	0	0	0	0
7	.	0	0	1	0	0
8	.	0	0	1	1	0
9	3	0	0	1	1	1
F_g		4	2	7	8	9
$\widetilde{F}_g$		4	4	3	3	3
$\widetilde{F}_g^b$		2	.	3	3	3
λ_g		2	.	2	1	0

Table 1: Example of a (reshaped) panel dataset with $G = 5$, $k = 3$, $\tau = 3$, $T = 9$ and $\widetilde{T} = 3$. Each column denotes the treatment path of a group. Years with non-missing outcome Y are shaded. corresponding values of F_g, $\widetilde{F}_g$, $\widetilde{F}_g^b$ and λ_g are reported in the bottom section of the table for each group. $\widetilde{F}_g^b$ denotes $\widetilde{F}_g$ under Design Restriction 1, while λ_g is computed under Design Restrictions 1 and 2. As a result, they are not defined for group 2, since Design Restriction 2 rules out groups such that $F_g \leq \tau$.

be the first period in the collapsed data where the treatment of group g changes.

The distinction above is practically relevant, since statistical softwares generally drop observations when variables of interest are missing. As a result, the first treatment change detected in the data used for the estimation procedure is $\widetilde{F}_g$, which does not always map to F_g. From the second equality in Equation 3, we see that $p(\widetilde{F}_g) = F_g$ if group g's treatment changes in a period where $Y_{g,t}$ is not missing. However, this is the only case where one can accurately detect switching groups from collapsed data.

Take the case of group g such that $D_{g,t} = D_{g,\tau}$ for $t \neq \tau + 1$ and $D_{g,\tau+1} \neq D_{g,\tau}$. In other words, group g's treatment changes only once at $t = \tau + 1$ and it goes back to the status quo level the period after. As a result, $F_g = \tau + 1$. However, $[(\tau+1) - \tau] \bmod k = 1$, hence we do not observe the treatment at $\tau + 1$ in the collapsed data. This implies that, since $\widetilde{D}_{g,\tilde{t}} = D_{g,\tau}$ for all $\tilde{t}$ in $\{1, ..., \widetilde{T}\}$, group g is wrongly categorized as a never-switcher if the treatment switch is detected from collapsed data. For instance, group 1 from Table 1 would be detected as a never-switcher with baseline treatment equal to 0, even though it takes on treatment 1 at period 4.

A slight modification of the main treatment solves this misclassification issue.

Design Restriction 1. *Instead of $D_{g,t}$, we use $D^b_{g,t} = 1\{t > F_g\}$ as the main treatment.*

By construction, the first period where $D^b_{g,t} \neq D^b_{g,t-1}$ is still equal to F_g. However, $G \times \widetilde{T}$ counterpart of $D^b_{g,t}$ takes 1 for all $t > F_g$ that satisfy Equation 1. As a result, on-and-off switchers, e.g. group 1 from Table 1, are correctly categorized as switchers even in collapsed data. Notice that $D^b_{g,t}$ is a binary and absorbing treatment, but it can be constructed from any type of treatment, i.e. binary, discrete or continuous. Throughout the paper, whenever we impose Design Restriction 1, $\widetilde{F}_g$ is defined using $D^b_{g,t}$.

Another possible detection error may happen when the first treatment change occurs before the first period when the outcome is non-missing. In this case, even under Design Restriction 1, switchers may be erroneously detected as always treated. Group 2 in Table 1 provides an example of this potential issue. The next design restriction rules out this case.

Design Restriction 2. *$F_g > \tau$ for all g.*

Design Restriction 2 can also be written as $D_{g,t} = D_{g,\tau}$ for all g and $t \leq \tau$.

Additional notation for collapsed data. Let

$$\Gamma_\ell = \{g : F_g - 1 + \ell \leq T_g, [(F_g - 1 + \ell) - \tau] \bmod k = 0\} \tag{4}$$

denote the set of switchers whose outcome is non-missing in period $F_g - 1 + \ell$ and that still have a valid control at that period. Let

$$\lambda_g = p(\widetilde{F}_g) - F_g = \begin{cases} 0 & \text{if } g \in \Gamma_1 \\ k - [(F_g - \tau) \bmod k] & \text{if } g \notin \Gamma_1 \end{cases} \tag{5}$$

denote the distance between F_g and the nearest subsequent period where $Y_{g,t}$ is non-missing. By construction, $\lambda_g \in \{0, 1, ..., k-1\}$, where $\lambda_g = 0$ for groups switching in a non-missing-outcome period. For instance, take the case of groups 3, 4 and 5 in Table 1. Group 5 switches on a non-missing-Y period, hence F_g satisfies Equation 1 and $\lambda_g = 0$. Conversely, the F_g of groups 3 and 4 is 2 and 1 periods away from a non-missing-Y period, respectively.

Lemma 1. *For any choice of ℓ, if $g_1, g_2 \in \Gamma_\ell$, then $\lambda_{g_1} = \lambda_{g_2}$*

Proof. If $g_1, g_2 \in \Gamma_1$, it follows from Equation 5 that $\lambda_{g_1} = \lambda_{g_2} = 0$. Similarly, take any $g_1, g_2 \in \Gamma_\ell$ with $\ell \neq 1$. By Equation 4, $Y_{g_1, F_{g_1} - 1 + \ell}$ and $Y_{g_2, F_{g_2} - 1 + \ell}$ are non-missing. It follows from Equation 1 that

$$[(F_{g_1} - 1 + \ell) - \tau] \bmod k = [(F_{g_2} - 1 + \ell) - \tau] \bmod k = 0$$

which reduces to

$$\begin{aligned}
[(F_{g_1} - 1 + \ell) - \tau] \bmod k + (1 - \ell) \bmod k &= [(F_{g_2} - 1 + \ell) - \tau] \bmod k + (1 - \ell) \bmod k \\
[(F_{g_1} - \tau) \bmod k] \bmod k &= [(F_{g_2} - \tau) \bmod k] \bmod k \\
(F_{g_1} - \tau) \bmod k &= (F_{g_2} - \tau) \bmod k \\
k - [(F_{g_1} - \tau) \bmod k] &= k - [(F_{g_2} - \tau) \bmod k]
\end{aligned}$$

where, the first equivalence comes by adding $(\ell - 1) \bmod k$ to both sides, the second from the fact that $(A + B) \bmod k = (A \bmod k + B \bmod k) \bmod k$ for any integers A,B and $k \geq 1$, and the third from the fact that $(A \bmod k) \bmod k = A \bmod k$ for any integers A and $k \geq 1$. As a result, $\lambda_{g_1} = \lambda_{g_2}$. ∎

Intuitively, Γ_ℓ contains all the groups such that their F_g is $(\ell - 1) \bmod k$ periods away from a non-missing-Y period. However, Lemma 1 cannot be inverted. Even if two groups share the same λ_g, it could be that their F_g is not the same. For instance, take the case of groups 1 and 3 from Table 1, where λ_g is the same, while F_g differs. Specifically, groups 1 and 3 are be both contained in Γ_1, while Γ_2 only includes group 1. Corollary 1 summarizes this fact.

Corollary 1. *If $\lambda_{g_1} = \lambda_{g_2}$ and $F_{g_1} = F_{g_2}$, then $g_1, g_2 \in \Gamma_\ell$ for at least one ℓ.*

Lastly, Lemma 2 shows a useful fact that will be used in the next paragraph.

Lemma 2. $F_g - 1 \geq p(\widetilde{F}_g - 1)$

Proof.

$$\begin{aligned} F_g - 1 - p(\widetilde{F}_g - 1) &= F_g - 1 - \tau - k[(\widetilde{F}_g - 1) - 1] \\ &= k + F_g - [\tau + k(\widetilde{F}_g - 1)] - 1 \\ &= k + F_g - p(\widetilde{F}_g) - 1 \\ &= k - (\lambda_g + 1) \\ &\geq 0 \end{aligned}$$

The first equality comes from the definition of $p(.)$, the second from rearranging, the third again from the definition of $p(.)$ and the fourth from the definition of λ_g. ■

Non-normalized actual-versus-status-quo estimator with collapsed data. Estimators of δ_ℓ become biased when the outcome is observed less frequently than the outcome. The $G \times \widetilde{T}$ counterpart of the $DID_{g,\ell}$ estimator from de Chaisemartin and D'Haultfoeuille (2024) is

$$\widetilde{DID}_{g,\ell} = \widetilde{Y}_{g,\widetilde{F}_g-1+\ell} - \widetilde{Y}_{g,\widetilde{F}_g-1} - \frac{1}{\widetilde{N}^g_{\widetilde{F}_g-1+\ell}} \sum_{\substack{g':\widetilde{D}_{g',1}=\widetilde{D}_{g,1}, \\ \widetilde{F}_{g'}>\tilde{F}_g-1+\ell}} \widetilde{Y}_{g',\widetilde{F}_g-1+\ell} - \widetilde{Y}_{g',\widetilde{F}_g-1} \tag{6}$$

where $\widetilde{N}^g_{\tilde{t}} = \#\{g' : \widetilde{D}_{g',1} = \widetilde{D}_{g,1}, \widetilde{F}_{g'} > \tilde{t}\}$ is the natural extension of N^g_t. Lastly, let

$$\widetilde{\Delta}_{g,\ell} = \widetilde{Y}_{g,\widetilde{F}_g-1+\ell} - \widetilde{Y}_{g,\widetilde{F}_g-1}$$

be the long difference of the outcome in the collapse data ℓ periods after $\widetilde{F}_g - 1$. To see where the bias comes from, take the case of Table 1. For $\ell = 1$, $\widetilde{\Delta}_{1,1} = \widetilde{Y}_{1,2} - \widetilde{Y}_{1,1}$, $\widetilde{\Delta}_{2,1}$ is not defined and $\widetilde{\Delta}_{g,1} = \widetilde{Y}_{g,3} - \widetilde{Y}_{g,2}$ for all other groups. Each of these long differences captures a change in outcome for switchers, but for a different number of post-first-switch periods. Take only groups 3, 4 and 5. For them it holds that

$$\widetilde{\Delta}_{g,1} = \widetilde{Y}_{g,3} - \widetilde{Y}_{g,2} = Y_{g,p(3)} - Y_{g,p(2)} = Y_{g,9} - Y_{g,6}$$

However, period 9 is λ_g periods away from the first switch of each of these groups. This means that the difference above identifies the outcome change of groups that have not switched treatment *at the same* *time*. As a result, $\widetilde{\Delta}_{g,1}$ captures the outcome evolution of groups 3, 4 and 5, but 3, 2 and 1 periods after $F_g - 1$, respectively.

Due to the misspecification of the long differences, the average of $\widetilde{DID}_{g,\ell}$ does not identify the average treatment effect for the switchers ℓ periods after their first switch. A solution to this issue is to estimate δ_ℓ using only switchers with the same λ_g. From Lemma 1, this condition is satified by restricting the sample of switchers to Γ_ℓ. To this end, let

$$\widetilde{DID}_\ell = \frac{1}{\#\{g \in \Gamma_\ell\}} \sum_{g \in \Gamma_\ell} \widetilde{DID}_{g,\ell}. \tag{7}$$

denote the average of $\widetilde{DID}_{g,\ell}$ across $g \in \Gamma_\ell$.

Proposition 1. *Under Assumptions 1, 3 and 4 from de Chaisemartin and D'Haultfoeuille (2024) and Design Restrictions 1 and 2,*

$$\mathbb{E}\left[\widetilde{DID}_\ell | \boldsymbol{D}\right] = \delta_{1+\lambda_g+(\ell-1)k}.$$

Proof.

$$\begin{aligned}
\widetilde{DID}_{g,\ell} =& \widetilde{Y}_{g,\widetilde{F}_g-1+\ell} - \widetilde{Y}_{g,\widetilde{F}_g-1} - \frac{1}{\widetilde{N}^g_{\widetilde{F}_g-1+\ell}} \sum_{\substack{g': \widetilde{D}_{g',1}=\widetilde{D}_{g,1}, \\ \widetilde{F}'_g > \widetilde{F}_g-1+\ell}} \widetilde{Y}_{g',\widetilde{F}_g-1+\ell} - \widetilde{Y}_{g',\widetilde{F}_g-1} \\
=& Y_{g,p(\widetilde{F}_g-1+\ell)} - Y_{g,p(\widetilde{F}_g-1)} - \frac{1}{N^g_{p(\widetilde{F}_g-1+\ell)}} \sum_{\substack{g': D_{g',1}=D_{g,1}, \\ p(\widetilde{F}'_g) > p(\tilde{F}_g-1+\ell)}} Y_{g',p(\widetilde{F}_g-1+\ell)} - Y_{g',p(\widetilde{F}_g-1)} \\
=& \Bigg(Y_{g,p(\widetilde{F}_g-1+\ell)} - Y_{g,F_g-1} - \frac{1}{N^g_{p(\widetilde{F}_g-1+\ell)}} \sum_{\substack{g': D_{g',1}=D_{g,1}, \\ p(\widetilde{F}'_g) > p(\tilde{F}_g-1+\ell)}} Y_{g',p(\widetilde{F}_g-1+\ell)} - Y_{g',F_g-1} \Bigg) \\
&- \Bigg(Y_{g,p(\widetilde{F}_g-1)} - Y_{g,F_g-1} - \frac{1}{N^g_{p(\widetilde{F}_g-1+\ell)}} \sum_{\substack{g': D_{g',1}=D_{g,1}, \\ p(\widetilde{F}'_g) > p(\tilde{F}_g-1+\ell)}} Y_{g',p(\widetilde{F}_g-1)} - Y_{g',F_g-1} \Bigg)
\end{aligned} \tag{8}$$

The second equality comes from Equation 2, the fact that $p(1) = \tau$ and Design Restriction 2, specifically $D_{g',\tau} = D_{g',1}$. The third equality comes from adding and subtracting Y_{g,F_g-1} and Y_{g',F_g-1}, and rearranging the terms. The first term in (8) is a $DID_{g,\ell}$ estimator, with ℓ equal to

$$\begin{aligned}
p(\widetilde{F}_g - 1 + \ell) - (F_g - 1) &= \tau + [(\widetilde{F}_g - 1 + \ell) - 1]k - F_g + 1 \\
&= 1 + [\tau + (\widetilde{F}_g - 1)k] - F_g + (\ell - 1)k \\
&= 1 + p(\widetilde{F}_g) - F_g + (\ell - 1)k \\
&= 1 + \lambda_g + (\ell - 1)k
\end{aligned}$$

From Lemma 2, we know that $p(\widetilde{F}_g - 1) \leq F_g - 1$. Hence, the second term in (8) is either 0, if $p(\widetilde{F}_g - 1) =$

$F_g - 1$, or a $DID^{pl}_{g,\ell}$ estimator, with ℓ equal to $k - (\lambda_g + 1)$. As a result,

$$
\begin{aligned}
\mathbb{E}[\widetilde{DID}_\ell|\mathbf{D}] =& \mathbb{E}\left[\frac{1}{\#\{g\in\Gamma_\ell\}}\sum_{g\in\Gamma_\ell}\widetilde{DID}_{g,\ell}\middle|\mathbf{D}\right] \\
=& \mathbb{E}\left[\frac{1}{\#\{g\in\Gamma_\ell\}}\sum_{g\in\Gamma_\ell}DID_{g,1+\lambda_g+(\ell-1)k}\middle|\mathbf{D}\right] - \mathbb{E}\left[\frac{1}{\#\{g\in\Gamma_\ell\}}\sum_{g\in\Gamma_\ell}DID^{pl}_{g,k-(\lambda_g+1)}\middle|\mathbf{D}\right] \\
=& \mathbb{E}\left[DID_{1+\lambda_g+(\ell-1)k}|\mathbf{D}\right] - \mathbb{E}\left[DID^{pl}_{k-(\lambda_g+1)}|\mathbf{D}\right] \qquad (9)\\
=& \delta_{1+\lambda_g+(\ell-1)k}.
\end{aligned}
$$

■

Proposition 1 specifies under what conditions dynamic effects can be retrieved from collapsed data. By *retrieving* the ℓ-th effect, we mean finding ℓ' such that $\mathbb{E}[DID_\ell|\mathbf{D}] = \mathbb{E}[\widetilde{DID}_{\ell'}|\mathbf{D}]$. More specifically, as long as there is at least one switcher and one not-yet-switcher with the same baseline treatment having non-missing $Y_{g,1+\lambda_g+(\ell-1)k}$ per value of $\lambda_g \in \{0,...,k-1\}$, one can retrieve DID_ℓ from a $\widetilde{DID}_{\ell'}$ estimator. Moreover, $\widetilde{DID}_{\ell'}$ is just DID_ℓ computed on collapsed data and including switchers that are ℓ periods away from their first switch at $\tilde{t} = \widetilde{F}_g - 1 + \ell$, i.e. groups in $\Gamma_{\ell'}$. Combining these two observations together, one can retrieve the ℓ-th dynamic effect from collapsed data by computing $DID_{\ell'}$ on a sample of switchers having the same λ_g and not-yet-switchers, provided that λ_g and ℓ' satisfy

$$\ell = 1 + \lambda_g + (\ell' - 1)k \qquad (10)$$

For instance, the first dynamic effect can be retrieved by computing DID_1 on groups having $\lambda_g = 0$, i.e. having their first switch on a period where the outcome is observed.

Testing the parallel trends assumption with collapsed data. Let

$$\widetilde{DID}^{pl}_{g,\ell} = \widetilde{Y}_{g,\widetilde{F}_g-1-\ell} - \widetilde{Y}_{g,\widetilde{F}_g-1} - \frac{1}{\widetilde{N}^g_{\widetilde{F}_g-1-\ell}} \sum_{\substack{g':\widetilde{D}_{g',1}=\widetilde{D}_{g,1},\\ \widetilde{F}_{g'}>\tilde{F}_g-1+\ell}} \widetilde{Y}_{g',\widetilde{F}_g-1-\ell} - \widetilde{Y}_{g',\widetilde{F}_g-1} \qquad (11)$$

denote the placebo estimator that mimicks the construction of $\widetilde{DID}_{g,\ell}$. Let $\widetilde{DID}^{pl}_\ell$ be the placebo counterpart of $\widetilde{DID}_\ell$. The following proposition adapts the results of Proposition 1 to placebo estimators.

Proposition 2. *Under Assumptions 1, 3 and 4 from de Chaisemartin and D'Haultfoeuille (2024) and Design Restrictions 1 and 2,*

$$
\begin{aligned}
&(a) \quad \widetilde{DID}^{pl}_\ell = DID^{pl}_{(\ell+1)k-(\lambda_g+1)} - DID^{pl}_{k-(\lambda_g+1)} \\
&(b) \quad \mathbb{E}\left[\widetilde{DID}^{pl}_\ell \middle| \boldsymbol{D}\right] = 0.
\end{aligned}
$$

Proof. The proof is essentially the same as in Proposition 1. The only difference lies in the indexation of the first term. $\widetilde{DID}_\ell^{pl}$ compares the outcome evolution of switchers and not-yet-switchers between $p(\tilde{F}_g - 1 - \ell)$ and $F_g - 1$. This is a *backward* comparison, in that, from Lemma 2

$$F_g - 1 \geq p(\tilde{F}_g - 1) > p(\tilde{F}_g - 1 - \ell)$$

for $\ell > 0$. As a result, $\widetilde{DID}_\ell^{pl}$ is defined over a number of periods before the first treatment switch equal to

$$(F_g - 1) - p(\tilde{F}_g - 1 - \ell) = (\ell + 1)k - (\lambda_g + 1). \tag{12}$$

■

The reason for computing placebo estimator is to test for parallel trends and no anticipation assumptions. The ideal setting would be to retrieve placebos for each effect, in that the dynamic estimator for the treatment effect ℓ periods after the first switch is unbiased only if the parallel trends assumption holds for at least ℓ periods. Unfortunately, the following corollary shows that not all placebos can be retrieved.

Corollary 2. *DID_ℓ^{pl} can be retrieved from collapsed data only if $\ell \geq k$.*

Proof. Recall that $\lambda_g \in \{0, ..., k-1\}$ and $\ell > 0$. From Equation 12, the minimum time distance between $F_g - 1$ and any prior period over which a $\widetilde{DID}_\ell^{pl}$ estimator is defined is equal to

$$\min_{\lambda_g, \ell} \; (F_g - 1) - p(\widetilde{F}_g - 1 - \ell) = (\ell + 1)k - (\lambda_g + 1)\Big|_{\ell = 1, \lambda_g = k-1} = k.$$

■

More specifically, Corollary 2 implies that the first $(k-1)$ placebos cannot be computed with collapsed data. The *earliest* placebo is DID_k, comparing the $\widetilde{F}_g - 2$-to-$\widetilde{F}_g - 1$ outcome evolution of groups with $\lambda_g = k - 1$. This is the case of group 3 in Table 1. Group 3 also satifies $p(\widetilde{F}_g - 1) = F_g - 1$, and, from Lemma 2, $k = \lambda_g + 1$. Replacing $k = \lambda_g + 1$ in Proposition 2a, the second term in the decomposition collapes to zero. As a result, the placebo from the $G \times \widetilde{T}$ data is equivalent to a placebo estimator from $G \times T$ data, although for a different ℓ. This is the only condition under which we can reliably retrieve placebos from collapsed data, as the following corollary shows.

Corollary 3. *$\widetilde{DID}_\ell^{pl} = DID_{k\ell}^{pl}$ if and only if $\lambda_g = k - 1$ for all $g \in \Gamma_\ell$.*

Recall from Lemma 1 that λ_g is constant within Γ_ℓ. As a result, the ℓ-th placebo from collapsed data compares the $(F_g - 1 - k\ell)$-to-$(F_g - 1)$ switchers-vs-not-yet-switchers outcome evolution for all groups such that $k\ell$-th effect is defined.

For all other groups, the construction of the placebo estimators with collapsed data loses its symmetry with the corresponding construction of the dynamic effects. For groups in Γ_ℓ such that $\lambda_g < k - 1$, the decomposition in Proposition 2a does not reduce to a single placebo estimator. In this case, $\widetilde{DID}_\ell^{pl}$ is unbiased for the *difference* of two placebo estimators. For instance, take the case of group 4 from Table

1. The first term of first group-specific placebo estimator in collapsed data for this group compares the outcome at $\widetilde{F}_g - 2$ to $\widetilde{F}_g - 1$. This amounts to computing the 3-to-6 outcome evolution, although the period before the first treatment switch for this group, i.e. the *true* $F_g - 1$, occurs at period 7. Consequently, the 3-to-6 outcome evolution can be rearranged as the 3-to-7 outcome evolution minus the 6-to-7. In both cases, subtracting the correspondent outcome evolution of not-yet-treated yields a valid placebo estimator (i.e., the fourth and first placebos, respectively).

Hence, even if each placebo from collapsed data is reassigned to its correct indexation, as in the case of dynamic effects, the interpretation of these estimators changes (unless Corollary 3 holds). Specifically, testing for $\mathbb{E}[\widetilde{DID}^{pl}_\ell|\mathbf{D}] = 0$ when $\lambda_g < k-1$ is equivalent to test for the equality of two placebo estimators. If the parallel trends assumption holds, then the conditional expectations of both terms in the right hand side of Proposition 2 given $\mathbf{D}$ are equal to 0, hence $\mathbb{E}[\widetilde{DID}^{pl}_\ell|\mathbf{D}] = 0$. However, the same result could be obtained whenever

$$\mathbb{E}[Y_{g,F_g-1-[(\ell+1)k-(\lambda_g+1)]} - Y_{g,F_g-1}|\mathbf{D}] = \mathbb{E}[Y_{g,F_g-1-[k-(\lambda_g+1)]} - Y_{g,F_g-1}|\mathbf{D}] \neq 0$$

i.e., if the pre-treatment trends are equal, but not null. This in turn implies that $\mathbb{E}[\widetilde{DID}^{pl}_\ell|\mathbf{D}] = 0$, meaning that a significance test on these placebo estimator would fail to detect violation of parallel trends. Consequently, testing for the insignificance of placebo estimators comes with low power, in the light of the potential for the type II errors described above.

References

Clément de Chaisemartin and Xavier D'Haultfoeuille. Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics*, 2024.