Online Appendices

Bellégo, C., Benatia, D. and V. Dortet-Bernadet (2024) The Chained Difference-in-Differences. To appear in the Journal of Econometrics.

Appendix A collects all the proofs and Appendix B provides additional results for the empirical application.

A Mathematical Appendix

A.1 Simple setting

Proof 1 (Proof of Proposition 1) *In order to identify and estimate the $ATT(t)$ in this simple setting, we impose the following standard assumptions.*

Assumption 10 (Sampling) *For all $t = 1, ..., \mathcal{T}$, $\{Y_{it}, Y_{it+1}, D_{i1}, D_{i2}, ..., D_{i\mathcal{T}}\}_{i=1}^{n_t}$ is independent and identically distributed (iid) conditional on $S_{t,t+1} = 1$.*

Assumption 11 (Missing Completely At Random) *For all $t = 1, ..., \mathcal{T}$,*

$$Pr(S_t = 1|Y_1, ..., Y_{\mathcal{T}}, D_1, ..., D_{\mathcal{T}}) = Pr(S_t = 1). \tag{28}$$

Assumption 12 (Unconditional Parallel Trends) *For all $t = 2, ..., \mathcal{T}$,*

$$E\left[Y_t(0) - Y_{t-1}(0)|G = 1\right] = E\left[Y_t(0) - Y_{t-1}(0)|C = 1\right] \ a.s.. \tag{29}$$

Assumption 13 (Irreversibility of Treatment) *For all $t = 2, ..., \mathcal{T}$,*

$$D_{t-1} = 1 \text{ implies that } D_t = 1. \tag{30}$$

Assumption 14 (Existence of Treatment and Control Groups)

$$P(G = 1) = 1 - P(C = 1) \in (0, 1). \tag{31}$$

Substituting (4) and rearranging yields

$$
\begin{aligned}
\widehat{ATT_{CD}}(t) &= \sum_{\tau=1}^{t-1} \frac{1}{n} \sum_{i=1}^{n} \left\{ \widehat{w_{i\tau\tau+1}^{G}} \left(y_{i\tau+1} - y_{i\tau} \right) - \widehat{w_{i\tau\tau+1}^{C}} \left(y_{i\tau+1} - y_{i\tau} \right) \right\} \\
&= \sum_{\tau=1}^{t-1} \frac{1}{n} \sum_{i=1}^{n} \left\{ \widehat{w_{i\tau\tau+1}^{G}} \left(\delta_{\tau+1} - \delta_{\tau} + \beta_t + \varepsilon_{i\tau+1} - \varepsilon_{i\tau} \right) - \widehat{w_{i\tau\tau+1}^{C}} \left(\delta_{\tau+1} - \delta_{\tau} + \varepsilon_{i\tau+1} - \varepsilon_{i\tau} \right) \right\} \\
&= \sum_{\tau=1}^{t-1} \left[\frac{1}{n} \sum_{i=1}^{n} \widehat{w_{i\tau\tau+1}^{G}} \beta_{\tau+1} + \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{i\tau\tau+1}^{G}} - \widehat{w_{i\tau\tau+1}^{C}} \right) \left(\varepsilon_{i\tau+1} - \varepsilon_{i\tau} \right) \right] \\
&= \sum_{\tau=2}^{t} \beta_{\tau} + \sum_{\tau=1}^{t-1} \left[\frac{1}{n} \sum_{i=1}^{n} \widehat{w_{i\tau\tau+1}^{G}} \left(\varepsilon_{i\tau+1} - \varepsilon_{i\tau} \right) - \frac{1}{n} \sum_{i=1}^{n} \widehat{w_{i\tau\tau+1}^{C}} \left(\varepsilon_{i\tau+1} - \varepsilon_{i\tau} \right) \right],
\end{aligned}
$$

where the second equality follows from the fact that $w_{i\tau\tau+1}^{G} \neq 0$ *and* $w_{i\tau\tau+1}^{C} \neq 0$ *only if* $y_{i\tau+1} - y_{i\tau}$ *is observed, and the fourth and fifth equalities follow from* $\sum_{i=1}^{n} \widehat{w_{i\tau\tau+1}^{G}} = \sum_{i=1}^{n} \widehat{w_{i\tau\tau+1}^{C}} = 1$*. The second term in the final expression vanishes in expectations from Assumptions 10 and 12.*

The second estimator is given by

$$
\begin{aligned}
\widehat{ATT_{CS}}(t) =& \frac{1}{n} \sum_{i=1}^{n} \left\{ \left(\widehat{w_{it}^{G}} y_{it} - \widehat{w_{i1}^{G}} y_{i1} \right) - \left(\widehat{w_{it}^{C}} y_{it} - \widehat{w_{i1}^{C}} y_{i1} \right) \right\} \\
=& \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{it}^{G}} - \widehat{w_{it}^{C}} \right) y_{it} - \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{i1}^{G}} - \widehat{w_{i1}^{C}} \right) y_{i1} \\
=& \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{it}^{G}} - \widehat{w_{it}^{C}} \right) \left(\alpha_i + \delta_t + \sum_{\tau=2}^{t} \beta_{\tau} D_{i\tau} + \varepsilon_{it} \right) - \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{i1}^{G}} - \widehat{w_{i1}^{C}} \right) \left(\alpha_i + \delta_1 + \varepsilon_{i1} \right) \\
=& \frac{1}{n} \sum_{i=1}^{n} \widehat{w_{it}^{G}} \sum_{\tau=2}^{t} \beta_{\tau} + \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{it}^{G}} - \widehat{w_{it}^{C}} - \widehat{w_{i1}^{G}} + \widehat{w_{i1}^{C}} \right) \alpha_i + \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{it}^{G}} - \widehat{w_{it}^{C}} \right) \delta_t \\
& - \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{i1}^{G}} - \widehat{w_{i1}^{C}} \right) \delta_1 + \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{it}^{G}} - \widehat{w_{it}^{C}} \right) \varepsilon_{i1} - \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{i1}^{G}} - \widehat{w_{i1}^{C}} \right) \varepsilon_{i1} \\
=& \sum_{\tau=2}^{t} \beta_{\tau} + \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{it}^{G}} - \widehat{w_{it}^{C}} \right) \varepsilon_{it} - \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{i1}^{G}} - \widehat{w_{i1}^{C}} \right) \varepsilon_{i1} \\
& + \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{it}^{G}} - \widehat{w_{it}^{C}} \right) \alpha_i - \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{i1}^{G}} - \widehat{w_{i1}^{C}} \right) \alpha_i,
\end{aligned}
$$

where the second and third terms vanish in expectations under Assumption 12. Finally, the expectation of the last term equates zero under Assumption 11.

Following similar steps, the corresponding long DiD estimator would be given by

$$\widehat{ATT_{LD}}(t) = = \sum_{\tau=2}^{t} \beta_\tau + \left[\frac{1}{n} \sum_{i=1}^{n} \widehat{w_{i1t}^{G}} \left(\varepsilon_{it} - \varepsilon_{i1} \right) - \frac{1}{n} \sum_{i=1}^{n} \widehat{w_{i1t}^{C}} \left(\varepsilon_{it} - \varepsilon_{i1} \right) \right],$$

if balanced panel data was available.

We will prove the asymptotic normality of these estimators in the general framework in Theorem 2. In this proof, we only derive their asymptotic variance for the mentioned example.

$$\begin{aligned} E\left[n \left(\widehat{ATT_{CD}}(t) - ATT(t) \right)^2 \right] &= nE\left[\left(\sum_{\tau=1}^{t-1} \left[\frac{1}{n} \sum_{i=1}^{n} \left(\widehat{w_{i\tau\tau+1}^{G}} - \widehat{w_{i\tau\tau+1}^{C}} \right) \left(\varepsilon_{i\tau+1} - \varepsilon_{i\tau} \right) \right] \right)^2 \right] \\ &= \sum_{\tau=1}^{t-1} \frac{1}{n} \sum_{i=1}^{n} E\left[\left(\widehat{w_{i\tau\tau+1}^{G}} - \widehat{w_{i\tau\tau+1}^{C}} \right)^2 \left(\varepsilon_{i\tau+1} - \varepsilon_{i\tau} \right)^2 \right] \\ &= \sum_{\tau=1}^{t-1} \frac{1}{n} \sum_{i=1}^{n} E\left[\left(\widehat{w_{i\tau\tau+1}^{G}} - \widehat{w_{i\tau\tau+1}^{C}} \right)^2 \right] E\left[\left(\varepsilon_{i\tau+1} - \varepsilon_{i\tau} \right)^2 \right], \\ &= \sum_{\tau=1}^{t-1} E\left[\left(\widehat{w_{i\tau\tau+1}^{G}} - \widehat{w_{i\tau\tau+1}^{C}} \right)^2 \right] E\left[\left(\varepsilon_{i\tau+1} - \varepsilon_{i\tau} \right)^2 \right], \end{aligned}$$

where the third equality follows from independence of G and $\varepsilon_{i\tau+1} - \varepsilon_{i\tau}$. As $n \to \infty$, the weak law of large numbers implies $E\left[\widehat{w_{i\tau\tau+1}^{G}}^2 \right] \xrightarrow{p} \frac{1}{P(S_\tau S_{\tau+1} G = 1)}$ because $S_{itt+1} a_i \in 0, 1$ and $E[S_{itt+1} a_i S_{jtt+1} a_j] = 0$ for $i \neq j$. Therefore, the asymptotic variance for $n \to \infty$ is

$$\begin{aligned} E\left[n \left(\widehat{ATT_{CD}}(t) - ATT(t) \right)^2 \right] &= \sum_{\tau=1}^{t-1} \left[\frac{1}{qp} + \frac{1}{q(1-p)} \right] \left[(\rho-1)^2 \sigma_\varepsilon^2 + \sigma_\eta^2 \right] \\ &= \frac{(t-1)}{qp(1-p)} \frac{2(1-\rho)}{1-\rho^2} \sigma_\eta^2 \\ &= 2 \frac{(t-1)}{qp(1-p)} \frac{1}{1+\rho} \sigma_\eta^2, \end{aligned}$$

when assuming $P(S_{i\tau} S_{i\tau+1}) = q \in (0,1)$ for all τ, i and $P(G_i = 1) = p \in (0,1)$. Note that this variance is bounded below by $\frac{1}{qp(1-p)}(t-1)\sigma_\eta^2$ when $\rho = 1$, and bounded above by

$\frac{1}{qp(1-p)}2(t-1)\sigma_\eta^2$, *when* $\rho = 0$. *Also, having a complete panel implies* $q = 1$ *in this setting. Similarly, the variance of the second estimator can be developed into*

$$
\begin{aligned}
E\left[n\left(\widehat{ATT_{CS}}(t) - ATT(t)\right)^2\right] =& E\left[\left(\widehat{w_{it}^G} - \widehat{w_{it}^C}\right)^2\right]\sigma_{\varepsilon_t}^2 + E\left[\left(\widehat{w_{i1}^G} - \widehat{w_{i1}^C}\right)^2\right]\sigma_{\varepsilon_1}^2 \\
&+ E\left[\left(\widehat{w_{it}^G} - \widehat{w_{it}^C}\right)^2\alpha_i^2\right] + E\left[\left(\widehat{w_{i1}^G} - \widehat{w_{i1}^C}\right)^2\alpha_i^2\right] \\
&- E\left[\left(\widehat{w_{i1}^G} - \widehat{w_{i1}^C}\right)\alpha_i\right]E\left[\left(\widehat{w_{it}^G} - \widehat{w_{it}^C}\right)\alpha_i\right] \\
=& E\left[\left(\widehat{w_{it}^G} - \widehat{w_{it}^C}\right)^2\right]\sigma_{\varepsilon_t}^2 + E\left[\left(\widehat{w_{i1}^G} - \widehat{w_{i1}^C}\right)^2\right]\sigma_{\varepsilon_1}^2 \\
&+ 2E\left[\left(\widehat{w_{it}^G} - \widehat{w_{it}^C}\right)^2\right]\sigma_\alpha^2, \\
=& \frac{1}{qp(1-p)}(0.5\sigma_{\varepsilon_t}^2 + \sigma_{\varepsilon_1}^2 + \sigma_\alpha^2),
\end{aligned}
$$

where the second equality follows from the independence of outcomes and sampling as well as G *and* α_i. *As* $n \to \infty$, *we have* $E\left[\widehat{w_{i\tau}^G}^2\right] \xrightarrow{p} \frac{1}{P(S_\tau G=1)} = \frac{1}{P(S_\tau S_{\tau+1}G=1)+P(S_{\tau-1}S_\tau G=1)} = \frac{1}{2pq}$ *for* $1 < \tau < \mathcal{T}$, *and* $E\left[\widehat{w_{i\tau}^G}^2\right] \xrightarrow{p} \frac{1}{P(S_\tau G=1)} = \frac{1}{pq}$ *for* $\tau = 1$ *or* $\tau = \mathcal{T}$, *because we have two overlapping samples in each period except for the first and last. For* $t < \mathcal{T}$, *the asymptotic variance as* $n \to \infty$ *is thus*

$$
E\left[n\left(\widehat{ATT_{CS}}(t) - ATT(t)\right)^2\right] \xrightarrow{p} \frac{1}{qp(1-p)}\left(1.5\sigma_{\varepsilon_1}^2 + \sigma_\alpha^2 + (t-1)\sigma_\eta^2\right), \quad \text{for } \rho = 1,
$$

$$
E\left[n\left(\widehat{ATT_{CS}}(t) - ATT(t)\right)^2\right] \xrightarrow{p} \frac{1}{qp(1-p)}\left(\sigma_\alpha^2 + 1.5\frac{\sigma_\eta^2}{1-\rho^2}\right), \quad \text{for } \rho \in (0,1),
$$

and therefore

$$
E\left[n\left(\widehat{ATT_{CS}}(t) - ATT(t)\right)^2\right] \xrightarrow{p} \frac{1}{qp(1-p)}\left(\sigma_\alpha^2 + 1.5\sigma_\eta^2\right), \quad \text{for } \rho = 0.
$$

In the complete panel setting, the weights are slightly different, and similar computations give

$$
E\left[n\left(\widehat{ATT_{CS}}(t) - ATT(t)\right)^2\right] \xrightarrow{p} \frac{1}{p(1-p)}\left[(\rho^{t-1}-1)^2\sigma_{\varepsilon_1}^2 + \sigma_\eta^2\sum_{\tau=0}^{t-2}\rho^{2\tau} + 2\sigma_\alpha^2\right],
$$

Following similar steps, it is easy to show that the asymptotic variance of the long-DiD for $n \to \infty$ is

$$E\left[n\left(\widehat{ATT_{LD}}(t) - ATT(t)\right)^2\right] \xrightarrow{p} \frac{1}{qp(1-p)}\left[(\rho^{t-1}-1)^2\sigma^2_{\varepsilon_1} + \sigma^2_\eta \sum_{\tau=0}^{t-2} \rho^{2\tau}\right],$$

that is $\frac{1}{qp(1-p)}2\sigma^2_\eta$ if $\rho = 0$ and $\frac{1}{qp(1-p)}(t-1)\sigma^2_\eta$ if $\rho = 1$, assuming the rotating samples consist of the same individuals over the entire time horizon. Therefore, the chained DiD estimator achieves the minimum variance, i.e. that of the long DiD estimator when idiosyncratic errors follow a random walk. The chained DiD estimator is therefore an efficient estimator in both the balanced and unbalanced panel settings when errors follow a random walk.

Comparing CD and CS variance for $\rho \in (0,1)$, we have that CD has smaller variance if $2(t-1)\sigma^2_\eta/(1+\rho) \leq 1.5\sigma^2_\eta/(1-\rho^2) + \sigma^2_\alpha$. If $\rho = 0$, this condition becomes $\sigma^2_\eta/2 \leq \sigma^2_\alpha$ for $t = 2$, and $5\sigma^2_\eta/2 \leq \sigma^2_\alpha$ for $t = 6$.

A.2 General framework

A.2.1 Proofs for rotating panel setting

Proof 2 (Proof of Theorem 1) *This proof focuses on the identification of parameters in the general framework with a rotating panel structure. Let us define $ATT_X(g,\tau) = E[Y_\tau(1) - Y_\tau(0)|X, G_g = 1]$ to write its first-difference as*

$$\begin{aligned}
\Delta ATT_X(g,\tau) &= ATT_X(g,\tau) - ATT_X(g,\tau-1) \\
&= E[Y_\tau(1) - Y_\tau(0)|X, G_g = 1] - E[Y_{\tau-1}(1) - Y_{\tau-1}(0)|X, G_g = 1] \\
&= E[Y_\tau - Y_{\tau-1}|X, G_g = 1, S_{\tau,\tau-1} = 1] - E[Y_\tau - Y_{\tau-1}|X, C = 1, S_{\tau,\tau-1} = 1] \\
&= A_X(g,\tau) - B_X(g,\tau)
\end{aligned}$$

where the third equality follows from the conditional parallel trends assumption and the sampling independence. Proofs of Corollary 1] and 2] show how to adjust for sample selection on observables.

We can use the above expression to develop $ATT(g,t)$ into

$$\begin{aligned}
ATT(g,t) &= E\left(E[Y_t(1) - Y_t(0)|X, G_g = 1]|G_g = 1\right) \\
&= E\left(ATT_X(g,t)|G_g = 1\right) \\
&= E\left(\sum_{\tau=g}^{t} \Delta ATT_X(g,\tau)|G_g = 1\right) \\
&= \sum_{\tau=g}^{t} E\left(\Delta ATT_X(g,\tau)|G_g = 1, S_{\tau,\tau-1} = 1\right) \\
&= \sum_{\tau=g}^{t} E\left(A_X(g,\tau) - B_X(g,\tau)|G_g = 1, S_{\tau,\tau-1} = 1\right)
\end{aligned} \tag{32}$$

with

$$\begin{aligned}
E\left(A_X(g,\tau)|G_g = 1, S_{\tau,\tau-1} = 1\right) &= E\left(Y_\tau - Y_{\tau-1}|G_g = 1, S_{\tau,\tau-1} = 1\right) \\
&= E_M\big(\,(Y_\tau - Y_{\tau-1})\,\frac{G_g S_{\tau,\tau-1}}{E[G_g S_\tau S_{\tau-1}]}\big)
\end{aligned} \tag{33}$$

by the law of iterated expectations and the definition of F_M. Following the proofs of Theorem 1 and B.1 of Callaway and Sant'Anna (2021), the second term can be developed into[38]

$$\begin{aligned}
&E(B_X(g,\tau)|G_g = 1, S_{\tau,\tau-1} = 1) = E(E[Y_\tau - Y_{\tau-1}|X, C, S_{\tau,\tau-1}]|G_g = 1, S_{\tau,\tau-1} = 1) \\
&= E\left(E[\frac{C}{1 - P(G_g = 1|X, G_g + C, S_{\tau,\tau-1})}(Y_\tau - Y_{\tau-1})|X, G_g + C = 1, S_{\tau,\tau-1}]|G_g, S_{\tau,\tau-1}\right) \\
&= \frac{E\left(G_g S_{\tau,\tau-1} E[\frac{C}{1-P(G_g|X,G_g+C,S_{\tau,\tau-1})}(Y_\tau - Y_{\tau-1})|X, G_g + C, S_{\tau,\tau-1}]|G_g + C, S_{\tau,\tau-1}\right)}{P(G_g = 1|G_g + C, S_{\tau,\tau-1})},
\end{aligned} \tag{34}$$

[38]We alleviate notations by dropping $= 1$ from conditioning sets.

where using the definition of p_g yields

$$
\begin{aligned}
\dots &= \frac{E\left(G_g S_{\tau,\tau-1} E[\frac{C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X, G_g + C, S_{\tau,\tau-1}]|G_g + C, S_{\tau,\tau-1}\right)}{P(G_g = 1|G_g + C, S_{\tau,\tau-1})} \\
&= \frac{E\left(E[\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X, G_g + C, S_{\tau,\tau-1}]|G_g + C, S_{\tau,\tau-1}\right)}{E(G_g|G_g + C, S_{\tau,\tau-1})} \\
&= \frac{E\left((G_g + C)E[\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X, G_g + C, S_{\tau,\tau-1}]|S_{\tau,\tau-1}\right)}{E(G_g|G_g + C, S_{\tau,\tau-1})E((G_g + C)|S_{\tau,\tau-1})} \\
&= \frac{E\left((G_g + C)E[\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X, G_g + C, S_{\tau,\tau-1}]|S_{\tau,\tau-1}\right)}{E(G_g|S_{\tau,\tau-1})} \\
&= \frac{E\left(E[(G_g + C)|X, S_{\tau,\tau-1}]E[\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X, G_g + C, S_{\tau,\tau-1}]|S_{\tau,\tau-1}\right)}{E(G_g S_{\tau,\tau-1})} \\
&= \frac{E\left(E[\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X, S_{\tau,\tau-1}]|S_{\tau,\tau-1}\right)}{E(G_g|S_{\tau,\tau-1})} \\
&= \frac{E\left(\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|S_{\tau,\tau-1}\right)}{E(G_g|S_{\tau,\tau-1})} \\
&= \frac{E_M\left(\frac{p_g(X)CS_{\tau,\tau-1}}{1-p_g(X)}(Y_\tau - Y_{\tau-1})\right)}{E_M(G_g S_{\tau,\tau-1})}
\end{aligned}
\tag{35}
$$

with

$$
\begin{aligned}
E_M\left(\frac{p_g(X)CS_{\tau,\tau-1}}{1-p_g(X)}\right) &= E_M(\frac{E_M(G_g|X, G_g + C = 1)C}{E_M(C|X, G_g + C = 1)}S_{\tau,\tau-1}) \\
&= E_M(\frac{E_M(G_g|X)E_M(C|X)}{E_M(C|X)}S_{\tau,\tau-1}) \\
&= E_M(E_M(G_g|X)S_{\tau,\tau-1}) \\
&= E_M(G_g S_{\tau,\tau-1}).
\end{aligned}
\tag{36}
$$

Finally, the proof that identification is not guaranteed with the repeated cross-sectional estimator (cross-section DiD) presented in Appendix B of Callaway and Sant'Anna (2021) follows from taking a counterexample. Under Assumption 2, it is possible that $E[Y_t|X, C = 1, S_t = 1] = E[Y_t|X, C = 1]$ but $E[Y_t|X, G_g = 1, S_t = 1] = E[Y_t|X, G_g = 1] + \alpha t$ with $\alpha > 0$. Following the steps of the proof it is is easy to show that the cross-section DiD identifies

$$ATT(g, t) + \alpha(t - g + 1).$$

Proof 3 (Proof of Theorem 2) *This proof is adapted from Callaway and Sant'Anna (2021) 's Theorem 2. We proceed in 3 steps.*

Parametric propensity scores and notations. *First, we introduce additional notations and explain Assumption 5 in Callaway and Sant'Anna (2021) about the estimation of propensity scores.*[39] *Let $\mathcal{W}_i = (Y_{it}, Y_{it+1}, X_i, G_{i1}, G_{i2}, ..., , G_{i\mathcal{T}}, C_i)'$ denote the data for an individual i observed in t and $t+1$. Assumption 5 in Callaway and Sant'Anna (2021) assumes that the propensity scores, parametrized as $p_g(X_i) = \Lambda(X_i'\pi_g^0)$ with $\Lambda(\cdot)$ being a known function (logit or probit), can be parametrically estimated by maximum likelihood. We denote $\hat{p}_g(X_i) = \Lambda(X_i'\hat{\pi}_g)$ where $\hat{\pi}_g$ are estimated by ML, $\dot{p}_g = \partial p_g(u)/\partial u$, and $\dot{p}_g(X) = \dot{p}_g(X_i'\pi_g^0)$. Under this assumption, the estimated parameter $\hat{\pi}_g$ is asymptotically linear, i.e.,*

$$\sqrt{n}(\hat{\pi}_g - \pi_g^0) = \frac{1}{\sqrt{n}} \sum_i \xi_g^\pi(\mathcal{W}_i) + o_p(1),$$

where $\xi_g^\pi(\mathcal{W}_i)$ is defined in (3.1) is Callaway and Sant'Anna (2021) and does not depend on the sampling process since X is observed for all individuals.

Let us now define,

$$\psi_{gt}(\mathcal{W}_i) = \psi_{gt}^G(\mathcal{W}_i) + \psi_{gt}^G(\mathcal{W}_i), \tag{37}$$

where

$$\begin{aligned}
\psi_{gt}^G(\mathcal{W}_i) =& w_{it,t-1}^G(g) \left[(Y_{it} - Y_{it-1}) - E_M\left[w_{it,t-1}^G(g)(Y_{it} - Y_{it-1})\right]\right], \\
\psi_{gt}^C(\mathcal{W}_i) =& w_{it,t-1}^C(g, X) \left[(Y_{it} - Y_{it-1}) - E_M\left[w_{it,t-1}^C(g, X)(Y_{it} - Y_{it-1})\right]\right] + M_{gt}'\xi_g^\pi(\mathcal{W}_i),
\end{aligned}$$

and

$$M_{gt} = \frac{E_M\left[X(\frac{CS_tS_{t-1}}{1-p_g(X)})^2 \dot{p}_g(X)\left[(Y_{it} - Y_{it-1}) - E_M\left[w_{it,t-1}^C(g, X)(Y_{it} - Y_{it-1})\right]\right]\right]}{E_M[\frac{p_g(X)C}{1-p_g(X)}]}.$$

is a k dimensional vector, k being the number of covariates in X. Finally, let $\widehat{\Delta ATT}_{g\leq t}$ and $\Delta ATT_{g\leq t}$ denote the vectors of all $\widehat{\Delta ATT}(g,t)$ and $\Delta ATT(g,t)$ for any $2 \leq g \leq t \leq \mathcal{T}$. Similarly, the collection of ψ_{gt} across all periods and groups such that $g \leq t$ is denoted by $\Psi_{g\leq t}$.

[39] This is a standard assumption in the literature so it is not reproduced here.

Asymptotic result for ΔATT. *Second, we show the asymptotic result for ΔATT. Recall that*

$$\widehat{ATT}(g,t) = \sum_{\tau=g}^{t} \widehat{\Delta ATT}(g,\tau),$$

where

$$\begin{aligned}\widehat{\Delta ATT}(g,\tau) &= \hat{E}_M\left[\frac{G_g S_{\tau-1} S_\tau}{\hat{E}_M[G_g S_{\tau-1} S_\tau]}(Y_\tau - Y_{\tau-1})\right] - \hat{E}_M\left[\frac{\frac{p_g(X) C S_{\tau-1} S_\tau}{1-p_g(X)}}{\hat{E}_M\left[\frac{p_g(X) C S_{\tau-1} S_\tau}{1-p_g(X)}\right]}(Y_\tau - Y_{\tau-1})\right] \\ &= \widehat{\Delta ATT}_g(g,\tau) - \widehat{\Delta ATT}_C(g,\tau),\end{aligned}$$

where $\hat{E}$ denotes the empirical mean. We will show separately that, for all for all $2 \leq g \leq t \leq \mathcal{T}$,

$$\sqrt{n}\left(\widehat{\Delta ATT}_g(g,t) - \Delta ATT_g(g,t)\right) = \frac{1}{\sqrt{n}} \sum_i \psi_{gt}^G(\mathcal{W}_i) + o_p(1), \tag{38}$$

and

$$\sqrt{n}\left(\widehat{\Delta ATT}_C(g,t) - \Delta ATT_C(g,t)\right) = \frac{1}{\sqrt{n}} \sum_i \psi_{gt}^C(\mathcal{W}_i) + o_p(1), \tag{39}$$

which together implies

$$\sqrt{n}\left(\widehat{\Delta ATT}(g,t) - \Delta ATT(g,t)\right) = \frac{1}{\sqrt{n}} \sum_i \psi_{gt}(\mathcal{W}_i) + o_p(1) \tag{40}$$

and the asymptotic normality of $\sqrt{n}\left(\widehat{\Delta ATT}_{g\leq t} - \Delta ATT_{g\leq t}\right)$ follows from the multivariate central limit theorem.

First, we show (38). Let $\beta_g = E_M[G_g S_{\tau-1} S_\tau]$ and $\hat{\beta}_g = \hat{E}_M[G_g S_{\tau-1} S_\tau]$ and note that

$$\sqrt{n}\left(\hat{\beta}_g - \beta_g\right) = \frac{1}{\sqrt{n}} \sum_i \left(G_{ig} S_{i\tau-1} S_{i\tau} - E[G_g S_{\tau-1} S_\tau]\right) \xrightarrow{p} 0, \text{ as } n \to +\infty.$$

Then, for all $2 \leq g \leq t \leq \mathcal{T}$,

$$\begin{aligned}\sqrt{n}(\widehat{\Delta ATT}_g(g,t) - \Delta ATT_g(g,t)) &= \sqrt{n}\hat{E}_M\left[\frac{G_g S_{t,t-1}}{\hat{\beta}_g}(Y_t - Y_{t-1})\right] - \sqrt{n} E_M\left[\frac{G_g S_{t,t-1}}{\beta_g}(Y_t - Y_{t-1})\right] \\ &= \frac{\sqrt{n}}{\hat{\beta}_g}(\hat{E}_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right] - \frac{\hat{\beta}_g}{\beta_g} E_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right]) \\ &= \frac{\sqrt{n}}{\hat{\beta}_g}(\hat{E}_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right] - E_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right]),\end{aligned}$$

and by the continuous mapping theorem,

$$
\begin{aligned}
\sqrt{n}(\widehat{\Delta ATT}_g(g,t) - \Delta ATT_g(g,t)) = {}& \frac{\sqrt{n}}{\beta_g}(\hat{E}_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right] - E_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right]) \\
& - \sqrt{n}\left(\frac{1}{\beta_g} - \frac{1}{\hat{\beta}_g}\right) E_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right] + o_p(1) \\
={}& \frac{\sqrt{n}}{\beta_g}(\hat{E}_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right] - E_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right]) \\
& - \frac{\sqrt{n}\left(\hat{\beta}_g - \beta_g\right)}{\beta_g^2} E_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right] + o_p(1) \\
={}& \frac{\sqrt{n}}{\beta_g}(\hat{E}_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right] - \frac{\hat{\beta}_g}{\beta_g} E_M\left[G_g S_{t,t-1}(Y_t - Y_{t-1})\right]) + o_p(1) \\
={}& \frac{1}{\sqrt{n}} \sum_i G_{ig} \frac{S_{it,t-1}(Y_{it} - Y_{it-1}) - \Delta ATT(g,t)}{\beta_g} + o_p(1) \\
={}& \frac{1}{\sqrt{n}} \sum_i w^G_{it,t-1}(g)\left[(Y_{it} - Y_{it-1}) - E_M\left[w^G_{t,t-1}(g)(Y_t - Y_{it-1})\right]\right] + o_p(1) \\
={}& \frac{1}{\sqrt{n}} \sum_i \psi^G_{gt}(\mathcal{W}_i) + o_p(1),
\end{aligned}
$$

proving (38).

Let us now turn to (39)*. For an arbitrary function* g*, let*

$$
w_t(g) = \frac{g(X) C S_{t,t-1}}{1 - g(X)}
$$

and note that

$$
\begin{aligned}
\sqrt{n}(\widehat{\Delta ATT}_C(g,t) - \Delta ATT_C(g,t)) = \sqrt{n}\left(\hat{E}_M\left[\frac{w_t(\hat{p}_g)}{\hat{E}_M[w_t(\hat{p}_g)]}(Y_t - Y_{t-1})\right] - E_M\left[\frac{w_t(p_g)}{E_M[w_t(p_g)]}(Y_t - Y_{t-1})\right]\right) \\
&= \frac{\sqrt{n}}{\hat{E}_M[w_t(\hat{p}_g)]}\left(\hat{E}_M\left[w_t(\hat{p}_g)(Y_t - Y_{t-1})\right] - \frac{\hat{E}_M[w_t(\hat{p}_g)]}{E_M[w_t(p_g)]}E_M\left[w_t(p_g)(Y_t - Y_{t-1})\right]\right) \\
&= \frac{\sqrt{n}}{\hat{E}_M[w_t(\hat{p}_g)]}\left(\hat{E}_M\left[w_t(\hat{p}_g)(Y_t - Y_{t-1})\right] - E_M\left[w_t(p_g)(Y_t - Y_{t-1})\right]\right) \\
&\quad - \frac{E_M\left[w_t(p_g)(Y_t - Y_{t-1})\right]}{\hat{E}_M[w_t(\hat{p}_g)]E_M[w_t(p_g)]}\sqrt{n}(\hat{E}_M\left[w_t(\hat{p}_g)\right] - \hat{E}_M\left[w_t(p_g)\right]) \\
&= \frac{1}{\hat{E}_M[w_t(\hat{p}_g)]}\sqrt{n}A_n(\hat{p}_g) - \frac{\Delta ATT_C(g,t)}{\hat{E}_M[w_t(\hat{p}_g)]}\sqrt{n}B_n(\hat{p}_g) \\
&= \frac{1}{E_M[w_t(p_g)]}\sqrt{n}A_n(\hat{p}_g) - \frac{\Delta ATT_C(g,t)}{E_M[w_t(p_g)]}\sqrt{n}B_n(\hat{p}_g) + o_p(1),
\end{aligned}
$$

where the last equality follows directly from Assumption 5, which implies Lemma A.2 and Lemma A.3 in Callaway and Sant'Anna (2021). Applying the mean value theorem yields

$$
\begin{aligned}
A_n(\hat{p}_g) =& \hat{E}_M\left[w_t(p_g)(Y_t - Y_{t-1})\right] - E_M\left[w_t(p_g)(Y_t - Y_{t-1})\right] \\
&+ \hat{E}_M\left[X\frac{CS_{t,t-1}}{(1 - p_g(X;\overline{\pi}))^2}\dot{p}_g(X;\overline{\pi})(Y_t - Y_{t-1})\right]'\left(\hat{\pi}_g - \pi_g^0\right),
\end{aligned}
$$

where $\overline{\pi}$ is an intermediate point that satisfies $|\overline{\pi}_g \pi_g^0| \leq |\hat{\pi}_g \pi_g^0|$ a.s. Thus, by Assumption 5, the previously mentioned Lemmas, and the Classical Glivenko-Cantelli's theorem,

$$
\begin{aligned}
A_n(\hat{p}_g) =& \hat{E}_M\left[w_t(p_g)(Y_t - Y_{t-1})\right] - E_M\left[w_t(p_g)(Y_t - Y_{t-1})\right] \\
&+ \hat{E}_M\left[X\frac{CS_{t,t-1}}{(1 - p_g(X))^2}\dot{p}_g(X)(Y_t - Y_{t-1})\right]'\left(\hat{\pi}_g - \pi_g^0\right) + o_p(n^{-1/2}),
\end{aligned}
$$

and using the same reasoning we obtain

$$
\begin{aligned}
B_n(\hat{p}_g) =& \hat{E}_M\left[w_t(p_g)\right] - E_M\left[w_t(p_g)\right] \\
&+ \hat{E}_M\left[X\frac{CS_{t,t-1}}{(1 - p_g(X))^2}\dot{p}_g(X)\right]'\left(\hat{\pi}_g - \pi_g^0\right) + o_p(n^{-1/2}).
\end{aligned}
$$

Combining the above results and making use of the same Lemma yields (40) hence concludes the proof for ΔATT. The asymptotic covariance is given by $\Sigma_\Delta = E\left[\Psi_{g\leq\tau}(\mathcal{W}_i)\Psi_{g\leq\tau}(\mathcal{W}_i)'\right]$.

Asymptotic result for ATT. *Finally, by making use of (40) we have that*

$$\begin{aligned}\sqrt{n}\left(\widehat{ATT}(g,t) - ATT(g,t)\right) =& \sum_{\tau=g}^{t} \sqrt{n}\left(\widehat{\Delta ATT}(g,\tau) - \Delta ATT(g,\tau)\right) \\ =& \frac{1}{\sqrt{n}} \sum_i \left[\sum_{\tau=g}^{t} \psi_{g\tau}(\mathcal{W}_i)\right] + o_p(1) \\ &\xrightarrow{d} N(0,\Sigma),\end{aligned}$$

where $\Sigma = E_M\left[\Phi_{g\leq\tau}(\mathcal{W}_i)\Phi_{g\leq\tau}(\mathcal{W}_i)'\right]$ *with* $\Phi_{g\leq\tau}(\mathcal{W}_i) = \sum_{\tau=g}^{t} \Psi_{g\leq\tau}(\mathcal{W}_i)$.

Therefore, the influence function of the chained DiD estimator corresponds to the sum of influence functions of the short-term DiD estimator.

A.2.2 Bootstrap implementation for rotating panel setting

Bootstrapped confidence bands for $\widehat{ATT}(g,t)$. The algorithm is as follows:

1. Draw a vector of $\mathbf{V^b} = (V_1, ..., V_i, ..., V_n)'$, where V_i's are iid zero mean random variables with unit variance, such as Bernoulli random variables with $Pr(V = 1 - \kappa) = \kappa/\sqrt{5}$ with $\kappa = (\sqrt{5}+1)/2$ as suggested by Mammen (1993).

2. Compute the bootstrap draw $\widehat{ATT}^{\star b}_{g\leq t} = \widehat{ATT}_{g\leq t} + \hat{\Phi}_{g\leq\tau}\mathbf{V^b}$ where $\hat{\Phi}_{g\leq\tau}$ is a consistent estimator of $\Phi_{g\leq\tau}$ (see below).

3. Compute $\hat{R}^{\star b}(g,t) = \sqrt{n}(\widehat{ATT}^{\star b}(g,t) - \widehat{ATT}(g,t))$ for each element of the vector $\widehat{ATT}^{\star b}_{g\leq t}$.

4. Repeat steps 1-3 B times. Note: do not re-estimate propensity scores and parameters for each draw.

5. Compute the bootstrapped covariance for each (g,t) as $\hat{\Sigma}^{1/2}(g,t) = (q_{0.75}(g,t) - q_{0.25}(g,t))/(z_{0.75} - z_{0.25})$, where $q_p(g,t)$ is the p^{th} sample quantile of $\hat{R}^\star$ (across B draws) and $z(g,t)$ is the p^{th} quantile the standard normal distribution.

6. For each b, compute $\text{t-stat}^b_{g\leq t} = \max_{(g,t)} |\hat{R}^{\star b}(g,t)|\hat{\Sigma}^{-1/2}(g,t)$.

7. Construct $\hat{c}_{1-\alpha}$ as the empirical $(1-\alpha)$ quantile of the B boostrap draws of $\text{t-stat}^b_{g\leq t}$.

8. Construct the bootstrapped simultaneous confidence band for $ATT(g,t)$ as $\hat{C}(g,t) = [\widehat{ATT}(g,t) \pm \hat{c}_{1-\alpha}\hat{\Sigma}^{-1/2}(g,t)/\sqrt{n}]$.

This procedure requires to compute $\hat{\Phi}_{g\leq\tau}$, represented here as a $K \times n$ matrix, with n being the number of observations and $K = \frac{\mathcal{T}(\mathcal{T}-1)}{2}$ being the number of (g,t) element for any $2 \leq g \leq t \leq \mathcal{T}$. This is done as follows:

1. For every (g,t), compute the n-dimensional vector ψ_{gt} with i^{th} element defined as $\psi_{gt}(i) = \psi^G_{gt}(i) + \psi^C_{gt}(i)$, where

$$\begin{aligned}\psi^G_{gt}(i) =& w^G_{it,t-1}(g)\left[(Y_{it} - Y_{it-1}) - E_M\left[w^G_{t,t-1}(g)(Y_t - Y_{t-1})\right]\right],\\ \psi^C_{gt}(i) =& w^C_{it,t-1}(g,X)\left[(Y_{it} - Y_{it-1}) - E_M\left[w^C_{t,t-1}(g,X)(Y_t - Y_{t-1})\right]\right] + M'_{gt}\xi^\pi_g(i),\end{aligned}$$

   where M_{gt}, a k dimensional vector (k being the number of covariates in X), is defined as

$$M_{gt} = \frac{E\left[X(\frac{CS_tS_{t-1}}{1-p_g(X)})^2\dot{p}_g(X)\left[(Y_{it} - Y_{it-1}) - E\left[w^C_{t,t-1}(g,X)(Y_t - Y_{it-1})\right]\right]\right]}{E[\frac{p_g(X)C}{1-p_g(X)}]},$$

   with $\hat{p}_g(X_i) = \Lambda(X'_i\hat{\pi}_g)$ being the parametric propensity score for covariates X_i and $\dot{p}_g(X) = \partial\Lambda(X'_i\hat{\pi}_g)/\partial(X'_i\hat{\pi}_g)$. Furthermore, $\xi^\pi_g(i)$ is a k-dimensional vector for each observation i, and is given by

$$\xi^\pi_g(i) = E_M\left[\frac{(G_g + C)\dot{p}_g(X)^2}{p_g(X_i)(1-p_g(X_i))}XX'\right]^{-1} X_i\frac{(G_g + C)(G_g - p_g(X_i))\dot{p}_g(X_i)}{p_g(X_i)(1-p_g(X_i))}.$$

2. Compute $\phi_{gt} = \sum_{\tau=g}^{t}\psi_{g\leq\tau}$ for all $2 \leq g \leq t \leq \mathcal{T}$.

3. Concatenate all ϕ_{gt}'s into a $K \times n$ matrix $\Phi_{g\leq t}$.

A.2.3 Summary parameters for rotating panel setting

The group-time average treatment effect $ATT(g,t)$ consists of a building-block to study the dynamic effect of a treatment on different cohorts of treated individuals. In most applications, the main causal parameters of interest are not the $ATT(g,t)$ themselves but aggregate parameters of these building-blocks. In this section, we briefly mention the three main parameters of interest as proposed in Callaway and Sant'Anna (2021), and show how the asymptotic results and multiplier bootstrap adapt to our setting.

Selective timing. The causal effect of a policy on the cohort treated in g is given by

$$\theta_S(g) = \frac{1}{\mathcal{T} - g + 1} \sum_{t=g}^{\mathcal{T}} ATT(g,t),$$

and thus, an average causal effect across groups can be written as

$$\theta_S = \sum_{g=2}^{\mathcal{T}} \theta_S(g) Pr(G = g).$$

Dynamic treatment. In the presence of dynamic effects, the researcher may be interested in accounting for the length of exposure to the treatment. The causal effects of an exposure length $e \in \{0, 1, 2, ...\}$ across groups is defined as

$$\theta_D(e) = \sum_{g=2}^{\mathcal{T}} \sum_{t=g+e}^{\mathcal{T}} ATT(g,t) Pr(G = g | t = g + e),$$

and therefore an average across exposure lengths is given by

$$\theta_D = \frac{1}{\mathcal{T} - 1} \sum_{e=1}^{\mathcal{T}-1} \theta_D(e).$$

Calendar time. In some applications, the researcher may be interested in how treatment effects differ with calendar time. Let us consider

$$\theta_C(t) = \sum_{g=2}^{t} ATT(g,t) Pr(G = g | g \leq t),$$

and therefore an average across exposure lengths is given by

$$\theta_C = \frac{1}{\mathcal{T} - 1} \sum_{t=2}^{\mathcal{T}} \theta_C(t),$$

The difference between θ_S, θ_D and θ_C is that the second and third attribute more weight to the groups with, respectively, longer exposure lengths, and treated in the earliest periods.

The asymptotic results and bootstrap procedure directly apply to the summary parameters. The following corollary summarizes these results.

Corollary 3 *Under the Assumptions of Theorem 2, for all parameters θ defined above, including those indexed by some variable, we have*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma_\theta),$$

as $n \to \infty$, where Σ_θ is defined in the proof and the bootstrap procedure is defined.

Proof 4 (Proof of Corollary 3) *All summary parameters defined in the text can be generically written as*

$$\theta = \sum_{g=2}^{\mathcal{T}} \sum_{t=2}^{\mathcal{T}} w_{gt} ATT(g,t),$$

where w_{gt} are some random weights. Estimators can be defined as

$$\hat{\theta} = \sum_{g=2}^{\mathcal{T}} \sum_{t=2}^{\mathcal{T}} \hat{w}_{gt} \widehat{ATT}(g,t),$$

where estimated weights are such that

$$\sqrt{n}(\hat{w}_{gt} - w_{gt}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{gt}^{w}(\mathcal{W}_i) + o_p(1),$$

with first and second moments given by $E[\xi_{gt}^{w}(\mathcal{W})] = 0$ and $E[\xi_{gt}^{w}(\mathcal{W})\xi_{gt}^{w}(\mathcal{W})']$ finite and positive definite. This condition is satisfied by the sample analogs of weights appearing in the summary parameters θ's presented in the main text. The application of Theorem 2 yields

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta) =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l^{w}(\mathcal{W}_i) + o_p(1) \\ &\xrightarrow{d} N(0, E[l^{w}(\mathcal{W})^2])\end{aligned}$$

as $n \to \infty$, and where

$$l^{w}(\mathcal{W}_i) = \sum_{g=2}^{\mathcal{T}} \sum_{t=2}^{\mathcal{T}} \left\{ w_{gt} \sum_{\tau=g}^{t} \psi_{g\tau}(\mathcal{W}_i) + \xi_{gt}^{w}(\mathcal{W}_i) ATT(g,t) \right\},$$

for ψ_{gt} defined in (58) and ξ^w_{gt} correspond to the estimation errors of weights. The same bootstrap procedure can hence be used for $\hat{\theta}$ using a consistent estimate of the influence function l^w.

A.2.4 Bootstrap for summary parameters

All estimators of summary parameters defined in the text can be generically written as

$$\hat{\theta} = \sum_{g=2}^{\mathcal{T}} \sum_{t=2}^{\mathcal{T}} \hat{w}_{gt} \widehat{ATT}(g,t),$$

where the weights $\hat{w}_{gt}$'s are possibly random. In simple settings they are not. For instance, let us consider

$$\theta_s(h) = \sum_{g=2}^{T} \sum_{t=2}^{T} w_{gt} ATT(g,t),$$

where $w_{gt} = \frac{1}{T-g+1}$ for $t \geq g$ and $g = h$, and 0 otherwise. The algorithm is as follows:

1. Draw a vector of $\mathbf{V^b} = (V_1, ..., V_i, ..., V_n)'$, where V_i's are iid zero mean random variables with unit variance, such as Bernoulli random variables with $Pr(V = 1 - \kappa) = \kappa/\sqrt{5}$ with $\kappa = (\sqrt{5}+1)/2$ as suggested by Mammen (1993).

2. Compute the bootstrap draw $\hat{\theta}^{\star b} = \hat{\theta} + \hat{L}'\mathbf{V^b}$ where $\hat{L}$ is a consistent estimator of the n-dimensional vector L with i^{th} element given by

$$L(i) = \sum_{g=2}^{T} \sum_{t=2}^{T} w_{gt} \phi_{gt}(i),$$

   where $\phi_{gt}(i)$ is defined in the previous algorithm.

3. Compute $\hat{R}^{\star b} = \sqrt{n}(\hat{\theta}^{\star b} - \hat{\theta})$.

4. Repeat steps 1-3 B times.

5. Compute the bootstrapped covariance as $\hat{\Sigma}^{1/2} = (q_{0.75} - q_{0.25})/(z_{0.75} - z_{0.25})$, where q_p is the p^{th} sample quantile of $\hat{R}^{\star}$ (across B draws) and z is the p^{th} quantile of the standard normal distribution.

6. For each b, compute t-stat$^b_{g \leq t} = \max_{(g,t)} |\hat{R}^{\star b}(g,t)| \hat{\Sigma}^{-1/2}(g,t)$.

7. Construct $\hat{c}_{1-\alpha}$ as the empirical $(1-\alpha)$ quantile of the B boostrap draws t-stat^b.

8. Construct the bootstrapped confidence interval for θ as $\hat{C} = [\hat{\theta} \pm \hat{c}_{1-\alpha}\hat{\Sigma}^{-1/2}/\sqrt{n}]$.

If the weights w_{gt}'s are random, the influence function is changed to

$$L(i) = \sum_{g=2}^{T}\sum_{t=2}^{T} w_{gt}\phi_{gt}(i) + \gamma_{gt}^{w}(i)ATT(g,t),$$

where $\gamma_{gt}^{w}(i)$ is an error function. For example, let us consider

$$\theta_s = \sum_{g=2}^{T}\sum_{t=2}^{T} w_{gt}ATT(g,t),$$

where $w_{gt} = P(G=g)\frac{1}{T-g+1}$ for $t \geq g$, and 0 otherwise. Define $\hat{w}_{gt} = \frac{1}{T-g+1}\frac{1}{n}\sum_{i=1}^{n} G_i$ for $t \geq g$ and 0 otherwise. A consistent estimator of the error function is given by

$$\hat{\gamma}_{gt}^{w}(i) = \frac{1}{T-g+1}\left(G_i - \frac{1}{n}\sum_{i=1}^{n} G_i\right).$$

A.2.5 Not yet treated as the control group

In the previous model, we assumed the existence of a true control group, i.e. a group of individuals that are never treated. In many applications, this situation is not realistic. Instead, the researcher can use the individuals that are "not yet treated", that is treated in $g > t$ to define a control group. The extension to this setting being developed in length in Callaway and Sant'Anna (2021), we only explain how it applies to the chained DiD.

Most importantly, the parallel trend assumption is modified to

Assumption 15 (Conditional Parallel Trends) *For all $t = 2, ..., \mathcal{T}$, $g = 2, ..., \mathcal{T}$, such that $g \leq t$,*

$$E\left[Y_t(0) - Y_{t-1}(0)|X, G_g = 1\right] = E\left[Y_t(0) - Y_{t-1}(0)|X, D_t = 0\right] \;\; a.s.. \tag{41}$$

Following minor modifications to Theorem C.1. in Callaway and Sant'Anna (2021), using the "not yet treated" as the control group only changes the weight $w_{\tau\tau-1}^{C}(g,X)$ used in our Theorem 1 to

$$w^C_{\tau\tau-1}(g,X) = \frac{p_{g,t}(X)(1-D_t)S_{\tau,\tau-1}}{1-p_{g,t}(X)}/E_M[\frac{p_{g,t}(X)(1-D_t)S_{\tau,\tau-1}}{1-p_{g,t}(X)}].$$

We observe two changes. First, the binary variable C becomes $1-D_t$. Second, generalized propensity score is now also a function of t: $p_{g,t}(X) = P(G_g = 1|X, (G_g = 1 \cup D_t = 1))$. The propensity scores must hence be estimated for pairs (g,t) because the control group evolves through time. The asymptotic properties of the two-step estimator remain similar, with minor changes to the asymptotic covariance.

A.2.6 General missing data patterns

Under Assumption 8, the data generating process consists of random draws from the mixture distribution $F_M(y, y', g_1, ..., g_{\mathcal{T}}, c, s_1, ..., s_{\mathcal{T}}, x)$ defined as

$$\sum_{t=1}^{\mathcal{T}} \lambda_{t-k,t} F_{Y_{t-k},Y_t,G_1,...,G_{\mathcal{T}},C,X|S_{t-k,t}}(y_{t-k}, y_t, g_1, ..., g_{\mathcal{T}}, C, X|s_{t-k,t} = 1),$$

where $\lambda_{t-k,t} = P(S_{t-k,t} = 1)$ is the probability of being sampled in both t and $t-k$, y and y' denote the outcome y_{t-k} and y_t, respectively, for an individual sampled at $t-k$ and t. Again, expectations under the mixture distribution does not correspond to population expectations. This difference arises because of different sampling probabilities across time periods and because Assumption 9 does not preclude from some forms of dependence between the sampling process and the unobservable heterogeneity in Y_{it}.

Proof 5 (Proof of Theorem 3) *Let us define the vector of parameters* $\Theta = \mathbf{ATT}$ *that includes all* $\theta_\tau = ATT(\tau)$*, for all* $\tau > 1$ *since* $\theta_1 = ATT(1) = 0$ *by construction. The inverse problem in* (23) *corresponds to the set of moment equalities*

$$E_M\left[h_i(\mathcal{W}_i|\Theta)\right] = 0_{L_\Delta}, \tag{42}$$

where $h_i(\mathcal{W}_i|\Theta)$ *is a* L_Δ*-dimensional vector of which each element is defined by*

$$\left[w^G_{i\tau\tau-k}(g)\left(Y_{i\tau} - Y_{i\tau-k}\right) - w^C_{i\tau\tau-k}(g,X)\left(Y_{i\tau} - Y_{i\tau-k}\right) - \theta_\tau + \theta_{\tau-k}\right], \tag{43}$$

possibly for all $\tau \geq 2$, and $1 \leq k < \tau$, with the weights

$$w^G_{\tau\tau-k}(g) = \frac{G_g S_{\tau,\tau-k}}{E_M[G_g S_{\tau,\tau-k}]}$$

and

$$w^C_{\tau\tau-k}(g, X) = \frac{p_g(X) C S_{\tau,\tau-k}}{1 - p_g(X)} / E_M[\frac{p_g(X) C S_{\tau,\tau-k}}{1 - p_g(X)}].$$

The previous asymptotic results in Theorem 1 and 2 apply to (43) up to minor modifications under Assumptions 3 - 5, a standard assumption on the parametric estimates of the propensity scores (Assumption 5 in Callaway and Sant'Anna (2021) or 4.4 in Abadie (2005)), and Assumptions 8 and 9 so that we can safely assume that consistency and asymptotic normality holds for $\widehat{\boldsymbol{\Delta ATT}}$ as $n \to \infty$ with covariance Ω, which is defined later. Here, we focus on the aspects of the proofs which differ, namely the optimal combination of each "chain link" using GMM. The optimal GMM estimator consists in minimizing

$$\hat{E}_M \left[h_i(\mathcal{W}_i|\Theta)\right]' \Omega^{-1} \hat{E}_M \left[h_i(\mathcal{W}_i|\Theta)\right], \tag{44}$$

with respect to Θ, using the optimal weighting matrix Ω^{-1} which corresponds the inverse of the covariance of h_i (Hansen, 1982), hence that of $\boldsymbol{\Delta ATT}$. Let us rewrite this problem as

$$\max_{\boldsymbol{ATT}} \quad -(\widehat{\boldsymbol{\Delta ATT}} - W\boldsymbol{ATT})'\Omega^{-1}(\widehat{\boldsymbol{\Delta ATT}} - W\boldsymbol{ATT}), \tag{45}$$

then the first-order condition with respect to $\boldsymbol{ATT}$ is given by

$$-2(\widehat{\boldsymbol{\Delta ATT}}) - W\boldsymbol{ATT})'\Omega^{-1}W = 0, \tag{46}$$

which, in turn, leads to the proposed estimator:

$$\widehat{\boldsymbol{ATT}} = (W'\Omega^{-1}W)^{-1}W'\Omega^{-1}\widehat{\boldsymbol{\Delta ATT}}. \tag{47}$$

The necessary and sufficient rank condition for GMM identification in this linear setting is that the rank of $\Omega^{-1}W$ is equal to the number of columns (Newey and McFadden, 1994). This condition is satisfied if both the covariance matrix Ω and the weight matrix W are non-singular. Remark further that if W is not full row rank then some $ATT(g,t)$ are not identified by the collection of $\Delta_k ATT(g,t)$'s identified in the dataset at hand.

Proving consistency requires introducing standard assumptions for GMM estimators. We *assume that (i) Ω and the weight matrix W are non-singular; (ii) the true value Θ_0 lies within a compact set; and (iii) $E_M[\sup_\Theta ||h_i(\mathcal{W}_i|\Theta)||] < \infty$. In addition to our previous assumptions, applying Theorem 2.6 in Newey and McFadden (1994) yields the desired consistency result. Assuming further that (iv) Θ_0 lies in the interior of the compact set; (v) $E[||h_i(\mathcal{W}_i|\Theta_0)||^2,] < \infty$; (vi) $W'\Omega^{-1}W$ non-singular, then asymptotic normality follows from Theorem 3.4 in Newey and McFadden (1994).*[40]

Note that the two-step GMM estimator requires estimating Ω. We proceed as follows. For every (g,t,k), compute the n-dimensional vector ψ_{gtk} with i^{th} element defined as

$$\psi_{gtk}(i) = \psi^G_{gtk}(i) + \psi^C_{gtk}(i), \tag{48}$$

where

$$\begin{aligned}
\psi^G_{gtk}(i) =& w^G_{it,t-k}(g)\left[(Y_{it} - Y_{it-k}) - E_M\left[w^G_{t,t-k}(g)(Y_t - Y_{t-k})\right]\right], \\
\psi^C_{gtk}(i) =& w^C_{it,t-k}(g,X)\left[(Y_{it} - Y_{it-k}) - E_M\left[w^C_{t,t-k}(g,X)(Y_t - Y_{t-k})\right]\right] + M'_{gtk}\xi^\pi_g(i),
\end{aligned}$$

where M_{gtk}, a k dimensional vector (k being the number of covariates in X), is defined as

$$M_{gtk} = \frac{E\left[X(\frac{CS_tS_{t-k}}{1-p_g(X)})^2\dot{p}_g(X)\left[(Y_{it} - Y_{it-k}) - E\left[w^C_{t,t-k}(g,X)(Y_t - Y_{it-k})\right]\right]\right]}{E[\frac{p_g(X)C}{1-p_g(X)}]},$$

with $\hat{p}_g(X_i) = \Lambda(X_i'\hat{\pi}_g)$ being the parametric propensity score for covariates X_i and $\dot{p}_g(X) = \partial\Lambda(X_i'\hat{\pi}_g)/\partial(X_i'\hat{\pi}_g)$. Furthermore, $\xi^\pi_g(i)$ is a k-dimensional vector for each observation i, and is given by

$$\xi^\pi_g(i) = E_M\left[\frac{(G_g + C)\dot{p}_g(X)^2}{p_g(X_i)(1 - p_g(X_i))}XX'\right]^{-1} X_i\frac{(G_g + C)(G_g - p_g(X_i))\dot{p}_g(X_i)}{p_g(X_i)(1 - p_g(X_i))}.$$

Concatenate all ψ_{gtk}'s into a $L_\Delta \times n$ matrix Ψ, and compute $\hat{\Omega} = \hat{E}[\Psi(i)\Psi(i)']$.

Therefore, the asymptotic covariance of $\widehat{\boldsymbol{ATT}}$ is

$$\boldsymbol{\Sigma} = (W'\Omega^{-1}W)^{-1}, \tag{49}$$

and its corresponding influence function to be used in the bootstrap procedure detailed in

[40] All the other sufficient conditions used by these theorems are trivially satisfied in this linear setting.

Online Appendix A.2.2 is the empirical counterpart of

$$\Phi = (W'\Omega^{-1}W)^{-1}W'\Omega^{-1}\Psi. \tag{50}$$

Finally, it is easy to show that the choice of the optimal weighting Ω^{-1} *is the same if the objective is instead to minimize the variance of a linear transformation* $R'\boldsymbol{ATT}$*, where* R *is a vector of weights, like for all the summary parameters considered in Online Appendix A.2.3. In that case, the bootstrap for summary parameters in Online Appendix A.2.4 apply with the (general) influence defined as follows. Let the weights in* R *be random, the influence function is changed to*

$$L(i) = R'\Phi(i) + \gamma^w(i)'\boldsymbol{ATT},$$

where γ^w *is the error function that depends on the randomness of the weights, as defined in Online Appendix A.2.4.*

Proof 6 (Proof of Corollary 1) *This proof focuses on the identification of parameters in the general framework with a rotating panel structure. We show that how attrition models can be combined with our approach. Let us modify the definition from Theorem 1,* $ATT_X(g,\tau) = E[Y_\tau(1) - Y_\tau(0)|X, G_g = 1]$ *to write its first-difference as*

$$\begin{aligned}
\Delta ATT_X(g,\tau) &= ATT_X(g,\tau) - ATT_X(g,\tau-1) \\
&= E[Y_\tau - Y_{\tau-1}|X, G_g = 1] - E[Y_\tau - Y_{\tau-1}|C = 1] \\
&= A_X(g,\tau) - B_X(g,\tau)
\end{aligned}$$

where the second equality follows from the conditional parallel trends assumption. Again, we can use the above expression to develop $ATT(g,t)$ *into*

$$ATT(g,t) = \sum_{\tau=g}^{t} E\left(A_X(g,\tau) - B_X(g,\tau)|G_g = 1\right) \tag{51}$$

with

$$\begin{aligned}
E\left(A_X(g,\tau)|G_g = 1\right) &= E\left(Y_\tau - Y_{\tau-1}|G_g\right) \\
&= E\left(E[Y_\tau - Y_{\tau-1}|X, G_g]|G_g\right) \\
&= E\left(E[Y_\tau - Y_{\tau-1}|X, G_g, S_\tau S_{\tau-1}]|G_g\right) \\
&= E\left(E[\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X, G_g]}(Y_\tau - Y_{\tau-1})|X, G_g]|G_g\right) \\
&= E\left(\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X, G_g]}(Y_\tau - Y_{\tau-1})|G_g\right) \\
&= E_M\big(\,(Y_\tau - Y_{\tau-1})\,\frac{G_g S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X, G_g]E[G_g]}\big) \\
&= E_M\big(\frac{E[G_g S_\tau S_{\tau-1}]}{E[S_\tau S_{\tau-1}|X, G_g]E[G_g]}\,(Y_\tau - Y_{\tau-1})\,\frac{G_g S_\tau S_{\tau-1}}{E[G_g S_\tau S_{\tau-1}]}\big)
\end{aligned} \tag{52}$$

by the law of iterated expectations and the definition of F_M*. The third equality follows from the conditional mean independence between* $Y_\tau - Y_{\tau-1}$ *and* $S_\tau S_{\tau-1}$ *conditional on* X *and treatment assignment. Remark that if the conditional probability* $E[S_\tau S_{\tau-1}|X, G_g]$ *does not depend on* X*, then we are back to Theorem 1. Similarly, we obtain*

$$
\begin{aligned}
&E(B_X(g,\tau)|G_g=1) = E(E[Y_\tau - Y_{\tau-1}|X,C,S_{\tau,\tau-1}]|G_g) \\
&= E(E[\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X,C]}(Y_\tau - Y_{\tau-1})|X,C]|G_g) \\
&= E\left(E[\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X,C]}\frac{C}{1-P(G_g=1|X,G_g+C)}(Y_\tau - Y_{\tau-1})|X,G_g+C=1]|G_g\right) \\
&= \frac{E\left(G_g E[\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X,C]}\frac{C}{1-P(G_g|X,G_g+C)}(Y_\tau - Y_{\tau-1})|X,G_g+C]|G_g+C\right)}{P(G_g=1|G_g+C)},
\end{aligned}
\tag{53}
$$

where using the definition of p_g *yields*

$$
\begin{aligned}
... &= \frac{E\left(G_g E[\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X,C]}\frac{C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X,G_g+C]|G_g+C\right)}{P(G_g=1|G_g+C)}, \\
&= \frac{E\left(E[\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X,C]}\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X,G_g+C]|G_g+C\right)}{E(G_g|G_g+C)}, \\
&= \frac{E\left((G_g+C)E[\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X,C]}\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X,G_g+C]\right)}{E(G_g|G_g+C)E(G_g+C)}, \\
&= \frac{E\left((G_g+C)E[\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X,C]}\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X,G_g+C]\right)}{E(G_g)}, \\
&= \frac{E\left(E[(G_g+C)|X]E[\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X,C]}\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X,G_g+C]\right)}{E(G_g)} \\
&= \frac{E\left(E[\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X,C]}\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})|X]\right)}{E(G_g)} \\
&= \frac{E\left(\frac{S_\tau S_{\tau-1}}{E[S_\tau S_{\tau-1}|X,C]}\frac{p_g(X)C}{1-p_g(X)}(Y_\tau - Y_{\tau-1})\right)}{E(G_g)} \\
&= \frac{E\left(\frac{E(G_g S_\tau S_{\tau-1})}{E[S_\tau S_{\tau-1}|X,C]E(G_g)}\frac{p_g(X)CS_\tau S_{\tau-1}}{1-p_g(X)}(Y_\tau - Y_{\tau-1})\right)}{E(G_g S_\tau S_{\tau-1})}
\end{aligned}
\tag{54}
$$

Therefore, one can estimate an attrition model ex-ante for each treatment group and the control group to obtain $E[S_\tau S_{\tau-1}|X,G_g]$ *and* $E[S_\tau S_{\tau-1}|X,C]$*, then reweight all first differences in outcome by* $\frac{E(G_g S_\tau S_{\tau-1})}{E[S_\tau S_{\tau-1}|X,G_g]E(G_g)}$ *and* $\frac{E(G_g S_\tau S_{\tau-1})}{E[S_\tau S_{\tau-1}|X,C]E(G_g)}$ *before applying our chained DiD estimator. Notice that, in general, the attrition model does not require to use the same* *explanatory variable X than the propensity score $p_g(X)$. Remark, however, that we must be able to identify $E(G_g)$ in the population.*

Proof 7 (Asymptotics for Corollary 1) *We now derive the asymptotic distributions considering the added uncertainty introduced by the estimation of a first-step selection model (propensity scores) under the assumptions of Corollary 1.*

First, we introduce additional notations about the estimation of propensity scores in the modified weights

$$\tilde{w}^G_{\tau\tau-1}(g,X) = \frac{E(G_g S_\tau S_{\tau-1})}{E[S_\tau S_{\tau-1}|X,G_g]E(G_g)} w^G_{\tau\tau-1}(g,X), \tag{55}$$

$$\tilde{w}^C_{\tau\tau-1}(g,X) = \frac{E(G_g S_\tau S_{\tau-1})}{E[S_\tau S_{\tau-1}|X,C]E(G_g)} w^C_{\tau\tau-1}(g,X). \tag{56}$$

Let us denote the added term

$$l_t(q_a) = \frac{E(G_g S_\tau S_{\tau-1})}{q_a(X)E(G_g)}, \tag{57}$$

with $q_a(X_i) = E[S_\tau S_{\tau-1}|X_i, G_g] = \Lambda(X_i'\rho_a^0)$ a parametric propensity scores with $\Lambda(\cdot)$ being a known function (logit or probit), that can be parametrically estimated by maximum likelihood. Remark that we could use other X's here. We denote $\hat{q}_a(X_i) = \Lambda(X_i'\hat{\rho}_g)$ where $\hat{\rho}_a$ are estimated by ML, $\dot{q}_a = \partial q_a(u)/\partial u$, and $\dot{q}_a(X) = \dot{q}_a(X_i'\rho_a^0)$. Under this assumption, the estimated parameter $\hat{\rho}_a$ is asymptotically linear.

Let us now define,

$$\psi_{gt}(\mathcal{W}_i) = \psi^G_{gt}(\mathcal{W}_i) + \psi^G_{gt}(\mathcal{W}_i), \tag{58}$$

where

$$\begin{aligned}
\psi^G_{gt}(\mathcal{W}_i) =& \tilde{w}^G_{it,t-1}(g,X)\left[(Y_{it} - Y_{it-1}) - E_M\left[\tilde{w}^G_{it,t-1}(g,X)(Y_{it} - Y_{it-1})\right]\right] + N^{G'}_{gt}\xi^\rho_g(\mathcal{W}_i),\\
\psi^C_{gt}(\mathcal{W}_i) =& \tilde{w}^C_{it,t-1}(g,X)\left[(Y_{it} - Y_{it-1}) - E_M\left[\tilde{w}^C_{it,t-1}(g,X)(Y_{it} - Y_{it-1})\right]\right] + M'_{gt}\xi^\pi_g(\mathcal{W}_i) + N^{C'}_{gt}\xi^\rho_C(\mathcal{W}_i),
\end{aligned}$$

with

$$M_{gt} = \frac{E_M\left[l_t(q_C)X(\frac{CS_tS_{t-1}}{1-p_g(X)})^2\dot{p}_g(X)\left[(Y_{it} - Y_{it-1}) - E_M\left[\tilde{w}^C_{it,t-1}(g,X)(Y_{it} - Y_{it-1})\right]\right]\right]}{E_M[\frac{p_g(X)C}{1-p_g(X)}]},$$

$$N_{gt}^C = -\frac{E_M\left[\frac{p_g(X)C}{1-p_g(X)} X \frac{E(G_g S_\tau S_{\tau-1})}{q_C(X;\overline{\pi})^2 E(G_g)} \dot{q}_C(X;\overline{\rho}) \left[(Y_{it}-Y_{it-1}) - E_M\left[\tilde{w}_{it,t-1}^C(g,X)(Y_{it}-Y_{it-1})\right]\right]\right]}{E_M[\frac{p_g(X)C}{1-p_g(X)}]},$$

$$N_{gt}^G = -\frac{E_M\left[G_g S_{t,t-1} X \frac{E(G_g S_\tau S_{\tau-1})}{q_C(X;\overline{\pi})^2 E(G_g)} \dot{q}_C(X;\overline{\rho}) \left[(Y_{it}-Y_{it-1}) - E_M\left[\tilde{w}_{it,t-1}^C(g,X)(Y_{it}-Y_{it-1})\right]\right]\right]}{E_M[G_g S_{t,t-1}]},$$

are a k dimensional vectors, k being the number of covariates in X. Finally, let $\widehat{\Delta ATT}_{g\leq t}$ and $\Delta ATT_{g\leq t}$ denote the vectors of all $\widehat{\Delta ATT}(g,t)$ and $\Delta ATT(g,t)$ for any $2 \leq g \leq t \leq \mathcal{T}$. Similarly, the collection of ψ_{gt} across all periods and groups such that $g \leq t$ is denoted by $\Psi_{g\leq t}$.

Second, we show the asymptotic result for ΔATT. Recall that

$$\widehat{ATT}(g,t) = \sum_{\tau=g}^{t} \widehat{\Delta ATT}(g,\tau),$$

where now

$$\begin{aligned}\widehat{\Delta ATT}(g,\tau) &= \hat{E}_M\left[l_t(q_g)\frac{G_g S_{\tau-1} S_\tau}{\hat{E}_M[G_g S_{\tau-1} S_\tau]}(Y_\tau - Y_{\tau-1})\right] - \hat{E}_M\left[l_t(q_C)\frac{\frac{p_g(X) C S_{\tau-1} S_\tau}{1-p_g(X)}}{\hat{E}_M\left[\frac{p_g(X) C S_{\tau-1} S_\tau}{1-p_g(X)}\right]}(Y_\tau - Y_{\tau-1})\right] \\ &= \widehat{\Delta ATT}_g(g,\tau) - \widehat{\Delta ATT}_C(g,\tau),\end{aligned}$$

Let us treat each term separately. For $\widehat{\Delta ATT}_C(g,\tau)$, take an arbitrary function g, let

$$w_t(g) = \frac{g(X) C S_{t,t-1}}{1-g(X)}$$

and note that

$$
\begin{aligned}
\sqrt{n}(\widehat{\Delta ATT}_C(g,t) - \Delta ATT_C(g,t)) &= \sqrt{n}\left(\hat{E}_M\left[\frac{l_t(\hat{q}_C)w_t(\hat{p}_g)}{\hat{E}_M[w_t(\hat{p}_g)]}(Y_t - Y_{t-1})\right] - E_M\left[\frac{l_t(q_C)w_t(p_g)}{E_M[w_t(p_g)]}(Y_t - Y_{t-1})\right]\right)\\
&= \frac{\sqrt{n}}{\hat{E}_M[w_t(\hat{p}_g)]}\left(\hat{E}_M\left[l_t(\hat{q}_C)w_t(\hat{p}_g)(Y_t - Y_{t-1})\right] - \frac{\hat{E}_M[w_t(\hat{p}_g)]}{E_M[w_t(p_g)]}E_M\left[l_t(q_C)w_t(p_g)(Y_t - Y_{t-1})\right]\right)\\
&= \frac{\sqrt{n}}{\hat{E}_M[w_t(\hat{p}_g)]}\left(\hat{E}_M\left[l_t(\hat{q}_C)w_t(\hat{p}_g)(Y_t - Y_{t-1})\right] - E_M\left[l_t(q_C)w_t(p_g)(Y_t - Y_{t-1})\right]\right)\\
&\quad - \frac{E_M\left[w_t(p_g)(Y_t - Y_{t-1})\right]}{\hat{E}_M[w_t(\hat{p}_g)]E_M[w_t(p_g)]}\sqrt{n}(\hat{E}_M\left[l_t(q_C)w_t(\hat{p}_g)\right] - \hat{E}_M\left[l_t(q_C)w_t(p_g)\right])\\
&= \frac{1}{\hat{E}_M[w_t(\hat{p}_g)]}\sqrt{n}A_n(\hat{p}_g,\hat{q}_C) - \frac{\Delta ATT_C(g,t)}{\hat{E}_M[w_t(\hat{p}_g)]}\sqrt{n}B_n(\hat{p}_g,\hat{q}_C)\\
&= \frac{1}{E_M[w_t(p_g)]}\sqrt{n}A_n(\hat{p}_g,\hat{q}_C) - \frac{\Delta ATT_C(g,t)}{E_M[w_t(p_g)]}\sqrt{n}B_n(\hat{p}_g,\hat{q}_C) + o_p(1),
\end{aligned}
$$

Applying the mean value theorem and the Classical Glivenko-Cantelli's theorem yield

$$
\begin{aligned}
A_n(\hat{p}_g,\hat{q}_C) =& \hat{E}_M\left[l_t(q_C)w_t(p_g)(Y_t - Y_{t-1})\right] - E_M\left[l_t(q_C)w_t(p_g)(Y_t - Y_{t-1})\right]\\
&+ \hat{E}_M\left[l_t(q_C)X\frac{CS_{t,t-1}}{(1-p_g(X;\overline{\pi}))^2}\dot{p}_g(X;\overline{\pi})(Y_t - Y_{t-1})\right]'\left(\hat{\pi}_g - \pi_g^0\right)\\
&- \hat{E}_M\left[w_t(p_g)X\frac{E(G_gS_\tau S_{\tau-1})}{q_C(X;\overline{\pi})^2E(G_g)}\dot{q}_C(X;\overline{\rho})(Y_t - Y_{t-1})\right]'\left(\hat{\rho}_C - \rho_C^0\right) + o_p(n^{-1/2}),
\end{aligned}
$$

$$
\begin{aligned}
B_n(\hat{p}_g) =& \hat{E}_M\left[l_t(q_C)w_t(p_g)\right] - E_M\left[l_t(q_C)w_t(p_g)\right]\\
&+ \hat{E}_M\left[l_t(q_C)X\frac{CS_{t,t-1}}{(1-p_g(X;\overline{\pi}))^2}\dot{p}_g(X;\overline{\pi})\right]'\left(\hat{\pi}_g - \pi_g^0\right)\\
&- \hat{E}_M\left[w_t(p_g)X\frac{E(G_gS_\tau S_{\tau-1})}{q_C(X;\overline{\pi})^2E(G_g)}\dot{q}_C(X;\overline{\rho})\right]'\left(\hat{\rho}_C - \rho_C^0\right) + o_p(n^{-1/2}),
\end{aligned}
$$

where $\overline{\pi}$ and $\overline{\rho}$ are intermediate points. Applying the same reasoning for $\Delta ATT_g(g,t)$ gives

$$
\sqrt{n}(\widehat{\Delta ATT}_g(g,t) - \Delta ATT_g(g,t)) = \frac{1}{E_M[w_t]}\sqrt{n}A_n(\hat{q}_C) - \frac{\Delta ATT_g(g,t)}{E_M[w_t]}\sqrt{n}B_n(\hat{p}_g,\hat{q}_C) + o_p(1),
$$

with

$$
w_t = G_gS_{t,t-1},
$$

and

$$
\begin{aligned}
A_n(\hat{q}_C) =& \hat{E}_M\left[l_t(q_g)w_t(Y_t - Y_{t-1})\right] - E_M\left[l_t(q_g)w_t(Y_t - Y_{t-1})\right] \\
&- \hat{E}_M\left[w_t X \frac{E(G_g S_\tau S_{\tau-1})}{q_g(X;\overline{\pi})^2 E(G_g)}\dot{q}_g(X;\overline{\rho})(Y_t - Y_{t-1})\right]'\left(\hat{\rho}_g - \rho_g^0\right) + o_p(n^{-1/2}),
\end{aligned}
$$

$$
\begin{aligned}
B_n(\hat{p}_g) =& \hat{E}_M\left[l_t(q_g)w_t\right] - E_M\left[l_t(q_g)w_t\right] \\
&- \hat{E}_M\left[w_t X \frac{E(G_g S_\tau S_{\tau-1})}{q_g(X;\overline{\pi})^2 E(G_g)}\dot{q}_g(X;\overline{\rho})\right]'\left(\hat{\rho}_g - \rho_g^0\right) + o_p(n^{-1/2}).
\end{aligned}
$$

Combining the above results and making use of the same Lemma yields (40) hence concludes the proof for ΔATT. The asymptotic covariance is given by $\Sigma_\Delta = E\left[\Psi_{g\leq\tau}(\mathcal{W}_i)\Psi_{g\leq\tau}(\mathcal{W}_i)'\right]$. The asymptotic result for ATT follows from the same reasoning as in our main theorem.

Proof 8 (Proof of Corollary 2) *This proof shows how to adapt our approach to a sampling assumption related to sequential missing at random (Hoonhout and Ridder, 2019). We focus on a simple case to illustrate the method. Let us assume that an individual can be sampled at most two periods in a row. The first time a unit is sampled is a function of X and G_g or C, but the probability that the unit is sampled again in the next period also depends on the realization of its outcome. Formally, we assume $Y_t \perp S_t | Y_{t-1}, X, G_g, S_t = 1$ but $Y_t \perp S_t | X, G_g, S_{t-1} = 0$. Let us only focus on $A_X(g,\tau)$ to show how the weights change. We have*

$$
\begin{aligned}
E\left(A_X(g,\tau)|G_g = 1\right) &= E\left(Y_\tau - Y_{\tau-1}|G_g\right) \\
&= E\left(E[Y_\tau - Y_{\tau-1}|X, G_g, Y_{\tau-1}]|G_g\right) \\
&= E\left(E[Y_\tau - Y_{\tau-1}|X, G_g, Y_{\tau-1}, S_\tau = 1, S_{\tau-1} = 1]|G_g\right) \\
&= E\left(E[\frac{S_\tau}{E[S_\tau|X, Y_{\tau-1}, S_{\tau-1} = 1, G_g]}(Y_\tau - Y_{\tau-1})|X, G_g, Y_{\tau-1}, S_{\tau-1}]|G_g\right) \\
&= E\left(E[\frac{S_\tau S_{\tau-1}}{E[S_\tau|X, Y_{\tau-1}, S_{\tau-1} = 1, G_g]E[S_{\tau-1}|X, G_g]}(Y_\tau - Y_{\tau-1})|X, G_g]|G_g\right) \\
&= E_M\big(\,(Y_\tau - Y_{\tau-1})\,\frac{G_g S_\tau S_{\tau-1}}{E[S_\tau|X, Y_{\tau-1}, S_{\tau-1} = 1, G_g]E[S_{\tau-1}|X, G_g]E[G_g]}\big) \\
&= E_M\big(\frac{E[G_g S_\tau S_{\tau-1}]}{E[S_\tau|X, Y_{\tau-1}, S_{\tau-1} = 1, G_g]E[S_{\tau-1}|X, G_g]E[G_g]}\,(Y_\tau - Y_{\tau-1})\,\frac{G_g S_\tau S_{\tau-1}}{E[G_g S_\tau S_{\tau-1}]}\big),
\end{aligned}
\tag{59}
$$

where the third equality follows from $Y_t \perp S_t | Y_{t-1}, X, G_g$. Remark that if the conditional probability $E[S_\tau|X, Y_{\tau-1}, S_{\tau-1}, G_g]$ does not depend on $Y_{\tau-1}, S_{\tau-1}$, then we are back to the *previous case.*

Therefore, our approach accommodates general attrition models with only minor modifications to the chained DiD estimator. We do not derive the asymptotic distribution for this case, but it follows from the same steps as in Corollary 1.

B Appendix for the Application

In this appendix, we present the detailed results in various tables. The next two tables give detailed results corresponding to Figures 1, 2, 3, and 4.

Table B.1: Effects on total workforce (exhaustively observed outcome)

	log(total workforce)				
	Long DiD	Chained DiD		Cross Section DiD	
	Exhaustive (1)	Exhaustive (2)	Unbalanced (3)	Exhaustive (4)	Unbalanced (5)
β_{-3}	-0.013	-0.013	0.032	-0.013	0.075
	[-0.084,0.058]	[-0.079,0.053]	[-0.059,0.123]	[-0.144,0.117]	[-0.609,0.76]
β_{-2}	-0.023	-0.023	-0.008	-0.023	0.012
	[-0.091,0.044]	[-0.087,0.041]	[-0.097,0.081]	[-0.13,0.083]	[-0.549,0.574]
β_{-1}	-0.005	-0.005	-0.005	-0.005	0.024
	[-0.054,0.044]	[-0.052,0.041]	[-0.066,0.056]	[-0.09,0.08]	[-0.442,0.491]
$ref.$	0	0	0	0	0
β_1	0.027	0.027	0.046	0.028	0.062
	[-0.026,0.08]	[-0.023,0.078]	[-0.021,0.112]	[-0.103,0.159]	[-0.372,0.495]
β_2	0.057**	0.056**	0.081**	0.057	0.092
	[-0.006,0.121]	[-0.005,0.118]	[0.002,0.159]	[-0.048,0.163]	[-0.426,0.61]
β_3	0.07**	0.068**	0.089**	0.069	0.051
	[0,0.14]	[0,0.135]	[0.006,0.172]	[-0.075,0.214]	[-0.506,0.607]
β_4	0.084**	0.077**	0.119***	0.083	0.087
	[0.002,0.167]	[0,0.154]	[0.024,0.214]	[-0.091,0.257]	[-0.53,0.704]
β_5	0.123***	0.111***	0.149***	0.119*	0.076
	[0.032,0.213]	[0.025,0.197]	[0.037,0.261]	[-0.032,0.269]	[-0.533,0.686]
Pre-trend	-0.014	-0.014	0.006	-0.014	0.037
	[-0.048,0.021]	[-0.049,0.021]	[-0.04,0.053]	[-0.079,0.051]	[-0.294,0.368]

Notes: This table shows the dynamic treatment effects relative to the beginning of the treatment obtained on a panel balanced on exhaustive variables. "Exhaustive" refers to the use of an exhaustively observed outcome without pretending that this variable is imperfectly observed. "Unbalanced" refers to the use of an exhaustively observed outcome pretending that this variable is observed from R&D survey, that is, from an unbalanced repeated panel. 95% confidence intervals are obtained from the multiplier bootstrap. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table B.2: Effects on highly qualified workers (exhaustively observed outcome)

	log(highly qualified workforce)				
	Long DiD	Chained DiD		Cross Section DiD	
	Exhaustive (1)	Exhaustive (2)	Unbalanced (3)	Exhaustive (4)	Unbalanced (5)
β_{-3}	0.033	0.033	0.088	0.033	0.194
	[-0.059,0.126]	[-0.053,0.12]	[-0.059,0.235]	[-0.081,0.148]	[-0.307,0.696]
β_{-2}	0.018	0.018	0.013	0.018	0.006
	[-0.066,0.103]	[-0.06,0.096]	[-0.131,0.157]	[-0.08,0.116]	[-0.411,0.423]
β_{-1}	0.018	0.018	0.015	0.018	0.019
	[-0.058,0.094]	[-0.054,0.09]	[-0.099,0.129]	[-0.065,0.101]	[-0.328,0.365]
$ref.$	0	0	0	0	0
β_1	0.022	0.022	0.02	0.024	0.034
	[-0.056,0.1]	[-0.048,0.091]	[-0.09,0.129]	[-0.083,0.13]	[-0.287,0.355]
β_2	0.044	0.043	0.067	0.045	0.065
	[-0.038,0.126]	[-0.032,0.118]	[-0.051,0.184]	[-0.05,0.141]	[-0.319,0.449]
β_3	0.075*	0.072**	0.061	0.072	0.02
	[-0.015,0.164]	[-0.008,0.152]	[-0.066,0.189]	[-0.05,0.194]	[-0.388,0.428]
β_4	0.109**	0.104**	0.108	0.105	0.008
	[0.007,0.211]	[0.011,0.196]	[-0.04,0.256]	[-0.04,0.25]	[-0.439,0.456]
β_5	0.125**	0.117***	0.128*	0.121*	0.069
	[0.014,0.236]	[0.016,0.217]	[-0.032,0.288]	[-0.014,0.257]	[-0.39,0.528]
Pre-trend	0.023	0.023	0.039	0.023	0.073
	[-0.022,0.068]	[-0.018,0.064]	[-0.035,0.113]	[-0.033,0.08]	[-0.173,0.319]

Notes: This table shows the dynamic treatment effects relative to the beginning of the treatment obtained on a panel balanced on exhaustive variables. “Exhaustive” refers to the use of an exhaustively observed outcome without pretending that this variable is imperfectly observed. “Unbalanced” refers to the use of an exhaustively observed outcome pretending that this variable is observed from R&D survey, that is, from an unbalanced repeated panel. 95% confidence intervals are obtained from the multiplier bootstrap. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: Effects on employment variables observed from R&D survey

	Unbalanced variables from R&D survey in log			
	Chained DiD		Cross Section DiD	
	total workforce (1)	researchers (2)	total workforce (3)	researchers (4)
β_{-3}	0.053	-0.041	0.104	0.106
	[-0.066,0.172]	[-0.194,0.112]	[-0.598,0.806]	[-0.249,0.462]
β_{-2}	0.012	0.017	0.064	0.056
	[-0.111,0.135]	[-0.103,0.137]	[-0.554,0.682]	[-0.246,0.357]
β_{-1}	0.009	0.007	0.02	0.033
	[-0.074,0.092]	[-0.104,0.119]	[-0.481,0.522]	[-0.219,0.285]
$ref.$	0	0	0	0
β_1	0.053	0.06	0.075	0.082
	[-0.032,0.139]	[-0.051,0.17]	[-0.409,0.56]	[-0.171,0.336]
β_2	0.038	0.124**	0.052	0.132
	[-0.073,0.149]	[0.001,0.248]	[-0.499,0.603]	[-0.164,0.427]
β_3	0.047	0.181***	0.057	0.139
	[-0.068,0.162]	[0.041,0.32]	[-0.525,0.638]	[-0.178,0.455]
β_4	0.129**	0.245***	0.081	0.148
	[0.007,0.25]	[0.093,0.397]	[-0.57,0.733]	[-0.189,0.485]
β_5	0.168***	0.288***	0.116	0.221
	[0.034,0.302]	[0.128,0.448]	[-0.546,0.778]	[-0.144,0.585]
Pre-trend	0.025	-0.006	0.063	0.065
	[-0.032,0.082]	[-0.075,0.064]	[-0.319,0.444]	[-0.116,0.246]

Notes: This table shows the dynamic treatment effects relative to the beginning of the treatment. The dynamic effects are estimated with outcome variables observed from R&D survey, that is, from an unbalanced repeated panel. 95% confidence intervals are obtained from the multiplier bootstrap. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

These estimates can be compared to the next results where administrative data are not reweighted using the exhaustively observed variables. Results are presented in Tables B.4, B.5, and B.6. This introduces some differences because there are more individuals in some periods compared to the results presented earlier.

Table B.4: Effects on total workforce (exhaustively observed outcome with the complete panel data)

	log(total workforce)				
	Long DiD	Chained DiD		Cross Section DiD	
	Exhaustive	Exhaustive	Unbalanced	Exhaustive	Unbalanced
	(1)	(2)	(3)	(4)	(5)
β_{-3}	-0.001	0.001	0.019	0	0.002
	[-0.056,0.054]	[-0.07,0.071]	[-0.067,0.105]	[-0.174,0.174]	[-0.46,0.464]
β_{-2}	-0.027	-0.027	-0.018	-0.022	0.001
	[-0.079,0.026]	[-0.092,0.038]	[-0.099,0.063]	[-0.173,0.128]	[-0.405,0.408]
β_{-1}	-0.017	-0.017	-0.013	-0.016	-0.022
	[-0.061,0.026]	[-0.068,0.033]	[-0.076,0.051]	[-0.14,0.107]	[-0.374,0.33]
$ref.$	0	0	0	0	0
β_1	0.045**	0.045**	0.051*	0.049	0.052
	[0.002,0.087]	[-0.006,0.095]	[-0.008,0.11]	[-0.073,0.17]	[-0.282,0.386]
β_2	0.066***	0.077***	0.088***	0.057	0.046
	[0.013,0.118]	[0.013,0.142]	[0.015,0.161]	[-0.086,0.199]	[-0.348,0.44]
β_3	0.069***	0.085***	0.089**	0.055	-0.001
	[0.009,0.128]	[0.01,0.159]	[0.004,0.174]	[-0.085,0.194]	[-0.408,0.406]
β_4	0.074***	0.085**	0.11***	0.05	-0.027
	[0.009,0.139]	[0,0.17]	[0.016,0.204]	[-0.117,0.216]	[-0.473,0.42]
β_5	0.094***	0.097**	0.134***	0.052	-0.07
	[0.019,0.169]	[0.003,0.19]	[0.028,0.24]	[-0.121,0.226]	[-0.545,0.406]
Pre-trend	-0.015	-0.015	-0.004	-0.013	-0.006
	[-0.043,0.013]	[-0.05,0.021]	[-0.047,0.039]	[-0.087,0.061]	[-0.238,0.225]

Notes: This table shows the dynamic treatment effects relative to the beginning of the treatment obtained on a panel balanced on exhaustive variables. "Exhaustive" refers to the use of an exhaustively observed outcome without pretending that this variable is imperfectly observed. "Unbalanced" refers to the use of an exhaustively observed outcome pretending that this variable is observed from R&D survey, that is, from an unbalanced repeated panel. 95% confidence intervals are obtained from the multiplier bootstrap. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table B.5: Effects on highly qualified workers (exhaustively observed outcome with the complete panel data)

	log(highly qualified workforce)				
	Long DiD	Chained DiD		Cross Section DiD	
	Exhaustive (1)	Exhaustive (2)	Unbalanced (3)	Exhaustive (4)	Unbalanced (5)
β_{-3}	0.019	0.025	0.07	0.04	0.102
	[-0.046,0.085]	[-0.053,0.102]	[-0.059,0.199]	[-0.091,0.171]	[-0.235,0.438]
β_{-2}	0.024	0.013	0.021	0.023	-0.026
	[-0.041,0.089]	[-0.062,0.087]	[-0.105,0.148]	[-0.09,0.137]	[-0.327,0.274]
β_{-1}	0.016	0.016	0.012	0.023	-0.017
	[-0.043,0.075]	[-0.053,0.085]	[-0.093,0.116]	[-0.076,0.122]	[-0.281,0.247]
$ref.$	0	0	0	0	0
β_1	0.028	0.028	0.02	0.031	0.019
	[-0.031,0.088]	[-0.04,0.095]	[-0.085,0.125]	[-0.069,0.132]	[-0.228,0.267]
β_2	0.045	0.047	0.077	0.044	0.044
	[-0.021,0.111]	[-0.028,0.123]	[-0.049,0.202]	[-0.068,0.155]	[-0.239,0.328]
β_3	0.078**	0.077**	0.064	0.072	-0.007
	[0.008,0.147]	[-0.006,0.16]	[-0.068,0.197]	[-0.041,0.185]	[-0.3,0.286]
β_4	0.116***	0.108***	0.101	0.103*	-0.052
	[0.042,0.19]	[0.017,0.199]	[-0.038,0.24]	[-0.03,0.235]	[-0.385,0.281]
β_5	0.134***	0.119***	0.13**	0.111*	-0.031
	[0.048,0.22]	[0.021,0.218]	[-0.023,0.283]	[-0.029,0.251]	[-0.382,0.32]
Pre-trend	0.02	0.018	0.034	0.029	0.019
	[-0.015,0.054]	[-0.022,0.058]	[-0.031,0.1]	[-0.032,0.089]	[-0.151,0.19]

Notes: This table shows the dynamic treatment effects relative to the beginning of the treatment obtained on a panel balanced on exhaustive variables. "Exhaustive" refers to the use of an exhaustively observed outcome without pretending that this variable is imperfectly observed. "Unbalanced" refers to the use of an exhaustively observed outcome pretending that this variable is observed from R&D survey, that is, from an unbalanced repeated panel. 95% confidence intervals are obtained from the multiplier bootstrap. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table B.6: Effects on employment variables observed from R&D survey with the complete panel data

	Unbalanced variables from R&D survey in log			
	Chained DiD		Cross Section DiD	
	total workforce (1)	researchers (2)	total workforce (3)	researchers (4)
β_{-3}	0.036	-0.03	0.017	0.081
	[-0.113,0.185]	[-0.149,0.089]	[-0.471,0.504]	[-0.148,0.31]
β_{-2}	0.028	-0.018	0.079	0.044
	[-0.114,0.171]	[-0.125,0.088]	[-0.368,0.526]	[-0.164,0.252]
β_{-1}	0.022	0.007	-0.016	0.037
	[-0.1,0.144]	[-0.082,0.097]	[-0.4,0.369]	[-0.149,0.223]
$ref.$	0	0	0	0
β_1	0.05	0.073	0.045	0.095
	[-0.066,0.165]	[-0.023,0.168]	[-0.32,0.409]	[-0.095,0.285]
β_2	0.047	0.152***	0.022	0.141
	[-0.084,0.177]	[0.048,0.257]	[-0.378,0.423]	[-0.059,0.341]
β_3	0.064	0.214***	0.014	0.134
	[-0.073,0.201]	[0.099,0.329]	[-0.402,0.43]	[-0.074,0.341]
β_4	0.12*	0.244***	-0.035	0.086
	[-0.029,0.269]	[0.121,0.368]	[-0.501,0.43]	[-0.137,0.309]
β_5	0.142**	0.297***	-0.046	0.138
	[-0.013,0.297]	[0.163,0.432]	[-0.508,0.415]	[-0.099,0.374]
Pre-trend	0.029	-0.014	0.027	0.054
	[-0.042,0.099]	[-0.073,0.046]	[-0.203,0.257]	[-0.065,0.173]

Notes: This table shows the dynamic treatment effects relative to the beginning of the treatment. The dynamic effects are estimated with outcome variables observed from R&D survey, that is, from an unbalanced repeated panel. 95% confidence intervals are obtained from the multiplier bootstrap. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$