Event study estimators computed by sdid_event

Diego Ciccia*

June 26, 2024

Adapting SDID to event study analysis. In what follows, I use the notation from Clarke et al. (2023) to present the estimation procedure for event-study Synthetic Difference-in-Differences (SDID) estimators. In a setting with N units observed over T periods, $N_{tr} < N$ units receive treatment D starting from period a, where $1 < a \leq T$. The treatment D is binary, i.e. $D \in \{0, 1\}$, and it affects some outcome of interest Y. The outcome and the treatment are observed for all (i, t) cells, meaning that the data has a balanced panel structure.

Henceforth, values of a are referred to as *cohorts* or *adoption periods*. The values of a are collected in A, i.e. the adoption date vector. For the sake of generality, we assume that $|A| > 1$, meaning that groups start receiving the treatment at different periods. The case with no differential timing can be simply retrieved by considering one cohort at a time. Time periods are indexed by $t \in \{1, ..., T\}$, while units are indexed by $i \in \{1, ...N\}$. Without loss of generality, the first $N_{co} = N - N_{tr}$ units are the never-treated group. As for the treated units, let I^a be the subset of $\{N_{co} + 1, ..., N\}$ containing the indices of units in cohort a. Lastly, we denote with N^a_{tr} and T^a_{tr} the number of units in cohort a and the number of periods from the the onset of the treatment in the same cohort to end of the panel, respectively. These two cohort-specific quantities can be aggregated into T_{post} from Clarke et al. (2023), i.e. the total number of post treatment periods of all the units in every cohort. Namely,

$$T_{post} = \sum_{a \in A} N^a_{tr} T^a_{tr} \tag{1}$$

is the sum of the products of N^a_{tr} and T^a_{tr} across all $a \in A$.

Disaggregating $\hat{\tau}^{sdid}_a$. The cohort-specific SDID estimator from Arkhangelsky et al. (2021) can be rearranged as follows:

$$\hat{\tau}^{sdid}_a = \frac{1}{T^a_{tr}} \sum_{t=a}^{T} \left(\frac{1}{N^a_{tr}} \sum_{i \in I^a} Y_{i,t} - \sum_{i=1}^{N_{co}} \omega_i Y_{i,t} \right) - \sum_{t=1}^{a-1} \left(\frac{1}{N^a_{tr}} \sum_{i \in I^a} \lambda_t Y_{i,t} - \sum_{i=1}^{N_{co}} \omega_i \lambda_t Y_{i,t} \right) \tag{2}$$

where λ_t and ω_i are the optimal weights chosen to best approximate the pre-treatment outcome evolution of treated and (synthetic) control units. τ^{sdid}_a compares the average outcome difference of treated in cohort a and never-treated before and after the onset of the treatment. In doing so, τ^{sdid}_a encompasses all the post-treatment periods. As a result, it is possible to estimate the treatment effect ℓ periods after the adoption of the treatment, with $\ell \in \{1, ..., T^a_{post}\}$, via a simple disaggregation of τ^{sdid}_a into the following event-study estimators:

*Sciences Po, Northwestern-Kellogg. diego.ciccia@kellogg.northwestern.edu

$$\hat{\tau}^{sdid}_{a,\ell} = \frac{1}{N^a_{tr}} \sum_{i \in I^a} Y_{i,a-1+\ell} - \sum_{i=1}^{N_{co}} \omega_i Y_{i,a-1+\ell} - \sum_{t=1}^{a-1} \left(\frac{1}{N^a_{tr}} \sum_{i \in I^a} \lambda_t Y_{i,t} - \sum_{i=1}^{N_{co}} \omega_i \lambda_t Y_{i,t} \right) \tag{3}$$

This estimator is very similar to those proposed by Borusyak et al. (2024), Liu et al. (2024) and Gardner (2022), when the design is a canonical DiD (de Chaisemartin and D'Haultfoeuille, 2023). The only difference lies in the fact that the outcomes are weighted via unit-time specific weights. Notice that by construction

$$\hat{\tau}^{sdid}_{a} = \frac{1}{T^a_{tr}} \sum_{\ell=1}^{T^a_{tr}} \hat{\tau}^{sdid}_{a,\ell} \tag{4}$$

that is, $\hat{\tau}^{sdid}_{a}$ is the sample average of the cohort-specific dynamic estimators $\hat{\tau}^{sdid}_{a,\ell}$.

Aggregating $\hat{\tau}^{sdid}_{a,\ell}$ estimators into event-study estimates. Let A_ℓ be the subset of cohorts in A such that $a-1+\ell \leq T$, i.e. such that their ℓ-th dynamic effect can be computed, and let

$$N^\ell_{tr} = \sum_{a \in A_\ell} N^a_{tr} \tag{5}$$

denote the number of units in cohorts where the ℓ-th dynamic effect can be estimated. We can use this notation to aggregate the cohort-specific dynamic effects into a single estimator. Let

$$\hat{\tau}^{sdid}_{\ell} = \sum_{a \in A_\ell} \frac{N^a_{tr}}{N^\ell_{tr}} \hat{\tau}^{sdid}_{a,\ell} \tag{6}$$

denote the weighted sum of the cohort-specific treatment effects ℓ periods after the onset of the treatment, with weights corresponding to the relative number of groups participating into each cohort. This estimator aggregates the cohort-specific treatment effects, upweighting more representative cohorts in terms of units included.

As in Equation 4, $\hat{\tau}^{sdid}_{\ell}$ can also be retrieved via disaggregation of another estimator from Clarke et al. (2023). Let $T_{tr} = \max_{a \in A} T^a_{tr}$ be the maximum number of post-treatment periods across all cohorts. Equivalently, T_{tr} can also be defined as the number of post-treatment periods of the earliest treated cohort. Then, one can show that

$$\widehat{ATT} = \frac{1}{T_{post}} \sum_{\ell=1}^{T_{tr}} N^\ell_{tr} \hat{\tau}^{sdid}_{\ell} \tag{7}$$

that is, the $\widehat{ATT}$ estimator from Clarke et al. (2023) is a weighted average of the event study estimators $\hat{\tau}^{sdid}_{\ell}$, with weights proportional to the number of units for which the ℓ-th effect can be computed.

Proof of Equation 7. Let T^a_{post} denote the total number of post-treatment periods across all units in cohort a.

$$\begin{aligned}
\widehat{ATT} &= \sum_{a\in A} \frac{T_{post}^a}{T_{post}} \hat{\tau}_a^{sdid} \\
&= \frac{1}{T_{post}} \sum_{a\in A} N_{tr}^a T_{tr}^a \hat{\tau}_a^{sdid} \\
&= \frac{1}{T_{post}} \sum_{a\in A} \sum_{\ell=1}^{T_{tr}^a} N_{tr}^a \hat{\tau}_{a,\ell}^{sdid} \\
&= \frac{1}{T_{post}} \sum_{\ell=1}^{T_{tr}} \sum_{a\in A_\ell} N_{tr}^a \hat{\tau}_{a,\ell}^{sdid} \\
&= \frac{1}{T_{post}} \sum_{\ell=1}^{T_{tr}} N_{tr}^\ell \hat{\tau}_\ell^{sdid}
\end{aligned}$$

where the first equality comes from the definition of $\widehat{ATT}$ in Clarke et al. (2023), the second equality from the definition of T_{post}^a, the third equality from Equation 4, the fourth equality from the fact that the sets $\{(a,\ell) : a \in A, 1 \le \ell \le T_{tr}^a\}$ and $\{(a,\ell) : 1 \le \ell \le T_{tr}, a \in A_\ell\}$ are equal and the fifth equality from the definition of $\hat{\tau}_\ell^{sdid}$.

References

Dmitry Arkhangelsky, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. Synthetic difference-in-differences. *American Economic Review*, 2021.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 2024.

Damian Clarke, Daniel Pailanir, Susan Athey, and Guido Imbens. Synthetic difference in differences estimation. *arXiv preprint*, 2023.

Clément de Chaisemartin and Xavier D'Haultfoeuille. Difference-in-differences for simple and complex natural experiments. *Working textbook under contract with Princeton University Press*, 2023.

John Gardner. Two-stage differences in differences. *arXiv preprint*, 2022.

Licheng Liu, Ye Wang, and Yiqing Xu. A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 2024.