Synthetic Difference-in-Differences with a Continuous Treatment

By Justin Waddy*

This paper presents a new estimation strategy for causal panel methods with a continuous treatment. Unlike the typical approach of difference-in-differences (DiD) with a continuous treatment, my estimator does not require the parallel trends assumption. The estimation strategy I propose uses synthetic difference-in-differences (SDID) to estimate treatment effects for a subpanel of units who share the same treatment dose, against a lowest-dose donor pool. Using Monte Carlo simulations calibrated to real-world data, I show the estimator performs well relative to DiD with a continuous treatment.

Difference-in-differences (DiD) with a continuous treatment allows researchers to study the effects of a newly introduced policy where units receive varying doses of treatment. Emerging literature has characterised the assumptions required to identify parameters of interest like a dose-outcome response (Callaway, Goodman-Bacon and Sant'Anna, 2024*a*). One of the strongest assumptions requires that in the absence of treatment, treated and untreated groups would have followed parallel trends over time. If researchers can find a methodology which relaxes this parallel trends assumption, treatment effects could be identified in more flexible setups, where outcomes trend in different directions before treatment.

I present an estimation strategy for causal estimators with a continuous treatment, which relaxes the parallel trends assumption. Using synthetic difference-in-differences (SDID) with groups of units who share the same treatment dose, my estimator creates a counterfactual for each cohort by using SDID to best match pre-trend outcomes via unit and time weights. One of the benefits of my estimation strategy is that it inherits the known properties of the SDID estimator, and is competitive with methods where typical DiD with a continuous treatment would have been used in the past.

My proposed estimator is similar to Callaway, Goodman-Bacon and Sant'Anna (2024*a*) who specify alternative estimation procedures to two-way fixed-effects (TWFE) for difference-in-differences with a continuous treatment. However, my estimator does not require parallel trends, and instead relies on assuming a latent factor model. Similar to Callaway, Goodman-Bacon and Sant'Anna (2024*a*), this paper has a 'sharp' treatment exposure which is continuous at a unit level, unlike the fuzzy setups in de Chaisemartin and d'Haultfœuille (2020). My estimator uses

* Waddy: University of Exeter (j.waddy@exeter.ac.uk). No external funding supported this research. There are no conflicts of interest.

functional parameters of interest which define a dose-response, such as the average causal response which was initially proposed by Angrist and Imbens (1995), and adapted for continuous treatment by Callaway, Goodman-Bacon and Sant'Anna (2024*a*). The estimation strategy I propose uses the usual SDID estimator from Arkhangelsky et al. (2021), and provides a method to extend it to continuous treatment. It resembles event study estimators insofar as it specifies separate SDID specifications for different cohorts of treated units, while holding baseline units fixed (Callaway and Sant'Anna, 2021; Callaway, Goodman-Bacon and Sant'Anna, 2024*b*). The methodology also uses synthetic control as a comparison in simulations, which was originally proposed by Abadie, Diamond and Hainmueller (2010).

In Section 1, I propose the estimator for the average treatment effect on the treated, define the potential outcomes framework and specify parameters of interest. In Section 2, I evaluate the finite-sample performance of the estimator via Monte Carlo simulations calibrated to data. Section 3 establishes the asymptotic properties of the estimator and provides formal results.

I. Estimator

Consider a balanced panel of N units indexed by $i \in \{1, \ldots, N\}$, observed over T periods indexed by $t \in \{1, \ldots, T\}$. Let $T_{\text{pre}} < T$ denote the last pre-treatment period, so that $T_{\text{post}} = T - T_{\text{pre}}$ periods follow the introduction of treatment. Units are treated after time T_{pre} after which each unit receives a time-invariant dose $D_i \in \mathcal{D} \subset \mathbb{R}_+$. Before treatment, each unit is assumed to have received the lowest treated dose in the panel $d_L = \min_i D_i$, which allows control units to have a dosage greater than zero as their baseline. The observed data are

$$\{(Y_{it}, D_i) : i = 1, \ldots, N;\ t = 1, \ldots, T\}.$$

Let $d_L < d_1 < d_2 < \cdots < d_C$ denote the distinct realised doses among treated units, such that C is the number of cohorts with a unique treatment dose greater than d_L. Let

$$\mathbf{C} := \{1, \ldots, C\}$$

index dose cohorts. For each $c \in \mathbf{C}$, let d_c denote the common realized dose of cohort c. Partition the sample into a baseline donor pool and C treated cohorts

$$\mathcal{G}_0 = \{i : D_i = d_L\}, \qquad \mathcal{G}_c = \{i : D_i = d_c\}.$$

Units in $\mathcal{G}_0$ receive the baseline dose d_L in every period. Units in $\mathcal{G}_c$ receive d_L for $t \leq T_{\text{pre}}$ and dose d_c for $t > T_{\text{pre}}$. Let $\mathcal{D}^+ = \{d_1, \ldots, d_C\}$ denote the support of doses above baseline. Define the treatment matrix as $\boldsymbol{W}_{it} = \mathbf{1}\{D_i > d_L, t > T_{\text{pre}}\}$. In terms of notation relating to cohorts, superscript (c) will denote vectors or matrices whose units are restricted or subsetted to cohort c, whereas subscript c will denote cohort-specific parameters, weights, and estimators.

For each cohort $c \in \mathbf{C}$, I estimate the treatment effect by applying SDID to the

sub-panel $\mathcal{G}_0 \cup \mathcal{G}_c$. Each cohort contains the same set of baseline-dose donor units, and cycles through the treated dose cohort. $\boldsymbol{W}$ and $\boldsymbol{Y}$ are subsetted to those units within the cohort, unit weights are computed, and then SDID is estimated for each panel. Figure 1 visualises how cohorts are rearranged into sub-panels. To compute

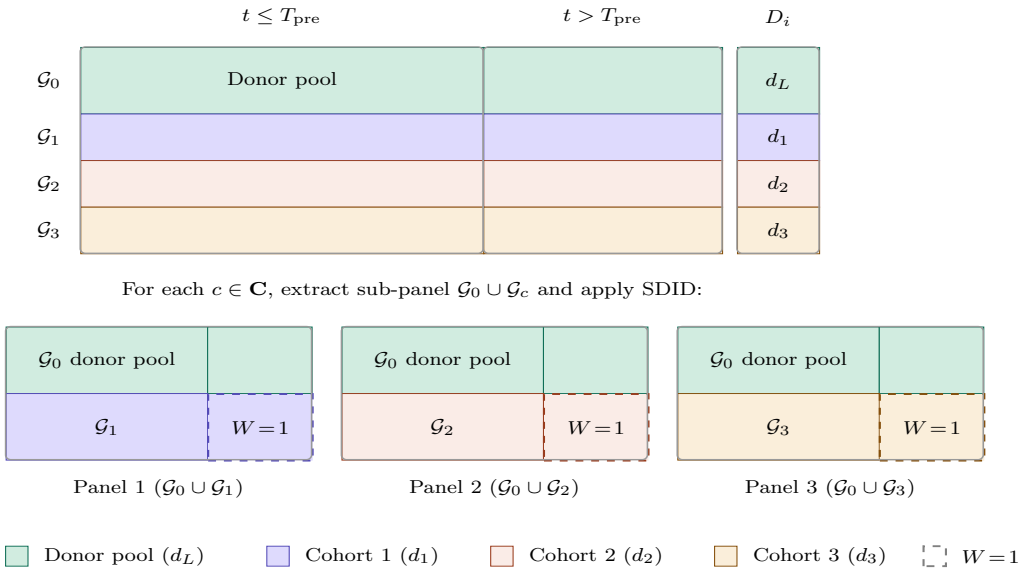


Figure 1. : Sub-panel construction for SDID with a continuous treatment. For each dose cohort $c \in \mathbf{C}$, a sub-panel $\mathcal{G}_0 \cup \mathcal{G}_c$ is formed and SDID is estimated. The donor pool $\mathcal{G}_0$ is held fixed across all sub-panels. The dashed region indicates post-treatment periods where $W_{it} = 1$ for treated units. The term in brackets of the legend denotes the post-treatment realised dose of units in a given cohort.

SDID, unit weights $\hat{\boldsymbol{\omega}}_c$ align pre-exposure trends for every sub-panel. Control units are the lowest-dose units in the panel such that the weighted combination of control units approximates the pre-treatment outcome path of treated cohort c,

$$\sum_{i \in \mathcal{G}_0} \hat{\omega}_{i,c} Y_{it} \;\approx\; \frac{1}{|\mathcal{G}_c|} \sum_{i \in \mathcal{G}_c} Y_{it}, \qquad \text{for all } t = 1, \ldots, T_{\text{pre}}.$$

Formally, to match the pretreatment outcomes of untreated units with the pre-treatment outcomes of treated units as described above, unit weights $\hat{\omega}_c$ solve the following optimisation problem for each cohort c. This is implemented by solving:

$$(\hat{\omega}_{0,c}, \hat{\omega}_c) = \underset{\omega_{0,c} \in \mathbb{R},\, \omega_c \in \Omega_c}{\arg\min}\; \ell_{\text{unit}}(\omega_{0,c}, \omega_c), \tag{1}$$

$$\ell_{\text{unit}}(\omega_{0,c}, \omega_c) = \sum_{t=1}^{T_{\text{pre}}} \left(\omega_{0,c} + \sum_{i=1}^{N_{\text{co}}} \omega_{i,c} Y_{it} - \frac{1}{|\mathcal{G}_c|} \sum_{i \in \mathcal{G}_c} Y_{it} \right)^2 + \zeta^2 T_{\text{pre}} \|\omega_c\|_2^2,$$

$$\Omega_c = \left\{ \omega_c \in \mathbb{R}^{N_{\text{co}}+|\mathcal{G}_c|} : \sum_{i=1}^{N_{\text{co}}} \omega_{i,c} = 1, \omega_{i,c} = |\mathcal{G}_c|^{-1} \text{ for all } i = N_{\text{co}}+1, \ldots, N_{\text{co}}+|\mathcal{G}_c| \right\}.$$

where

$$\zeta = (N_{\text{tr}} T_{\text{post}})^{1/4} \hat{\sigma}, \qquad \hat{\sigma}^2 = \frac{1}{N_{\text{co}}(T_{\text{pre}}-1)} \sum_{i=1}^{N_{\text{co}}} \sum_{t=1}^{T_{\text{pre}}-1} (\Delta_{it} - \bar{\Delta})^2, \tag{2}$$

This is the usual minimisation problem from Arkhangelsky et al. (2021), but for each cohort $c \in \mathbf{C}$. The regularization parameter is uniform across cohorts and matches the size of a typical one-period change in outcomes $Y_{i(t+1)} - Y_{it}$ for untreated units before treatment. Δ_{it} and $\bar{\Delta}$ are given by

$$\Delta_{it} = Y_{i(t+1)} - Y_{it}, \qquad \bar{\Delta} = \frac{1}{N_{\text{co}}(T_{\text{pre}}-1)} \sum_{i=1}^{N_{\text{co}}} \sum_{t=1}^{T_{\text{pre}}-1} \Delta_{it}.$$

Time weights are chosen as usual such that the weighted sum of pre-treatment outcomes for donor units approximates their average post-treatment outcome level,

$$\sum_{t=1}^{T_{\text{pre}}} \hat{\lambda}_t Y_{it} \approx \frac{1}{T_{\text{post}}} \sum_{t=T_{\text{pre}}+1}^{T} Y_{it},$$

Unlike unit weights, the time weights do not vary across cohorts because the set of control units $\mathcal{G}_0$ is common to all sub-panels. More formally, time weights λ solve:

$$(\hat{\lambda}_0, \hat{\lambda}) = \underset{\lambda_0 \in \mathbb{R},\, \lambda \in \Lambda}{\arg\min}\ \ell_{\text{time}}(\lambda_0, \lambda), \tag{3}$$

where

$$\ell_{\text{time}}(\lambda_0, \lambda) = \sum_{i=1}^{N_{\text{co}}} \left(\lambda_0 + \sum_{t=1}^{T_{\text{pre}}} \lambda_t Y_{it} - \frac{1}{T_{\text{post}}} \sum_{t=T_{\text{pre}}+1}^{T} Y_{it} \right)^2,$$

$$\Lambda = \left\{ \lambda \in \mathbb{R}_+^T : \sum_{t=1}^{T_{\text{pre}}} \lambda_t = 1,\ \lambda_t = T_{\text{post}}^{-1} \text{ for all } t = T_{\text{pre}}+1, \ldots, T \right\}.$$

Algorithm 1 specifies the estimation strategy for SDID with a continuous treatment.

The algorithm cycles through each treated cohort, computes unit weights, and computes the SDID estimator via a weighted DiD regression on the sub-panel. This algorithm can also be generalised to other estimators like SC and DiD, by using the relevant estimator on the repeated sub-panel. The difference between

Algorithm 1 SDID with a Continuous Treatment

1: Data: $\mathbf{Y}$, D_i, T_{pre}
2: Result: Cohort-specific estimate $\hat{\tau}_c$
3: Compute regularisation parameter ζ via (2);
4: Compute time weights $\hat{\lambda}$ via (3);
5: **for** $c \in \mathbf{C}$ **do**
6: Subset $\boldsymbol{Y}$ to units in $\mathcal{G}_0 \cup \mathcal{G}_c$, and compute unit weights $\hat{\omega}_c$ via (1)
7: Compute the SDID estimator via the weighted DiD regression:

$$\left(\hat{\tau}_c, \hat{\mu}_c, \hat{\boldsymbol{\alpha}}_c, \hat{\boldsymbol{\beta}}_c\right) = \underset{\tau,\mu,\alpha,\beta}{\arg\min} \left\{ \sum_{i \in \mathcal{G}_0 \cup \mathcal{G}_c} \sum_{t=1}^{T} (Y_{it} - \mu - \alpha_i - \beta_t - W_{it}^{(c)}\tau)^2 \hat{\omega}_{i,c}\, \hat{\lambda} \right\}$$

8: **end for**

average post-treatment and pre-treatment outcomes for the respective estimators can be written as

$$\hat{\delta}_{c,i}^{\text{did}} = \frac{1}{T_{\text{post}}} \sum_{t=T_{\text{pre}}+1}^{T} Y_{it} - \frac{1}{T_{\text{pre}}} \sum_{t=1}^{T_{\text{pre}}} Y_{it},$$

$$\hat{\delta}_{c,i}^{\text{sc}} = \frac{1}{T_{\text{post}}} \sum_{t=T_{\text{pre}}+1}^{T} Y_{it},$$

$$\hat{\delta}_{c,i}^{\text{sdid}} = \frac{1}{T_{\text{post}}} \sum_{t=T_{\text{pre}}+1}^{T} Y_{it} - \sum_{t=1}^{T_{\text{pre}}} \hat{\lambda}_t^{\text{sdid}} Y_{it}.$$

For each dose cohort $c \in \mathbf{C}$, the cohort-specific estimator is defined by the difference-in-differences between treated and untreated cohorts

$$\hat{\tau}_c(d_c) = \frac{1}{|\mathcal{G}_c|} \sum_{i \in \mathcal{G}_c} \hat{\delta}_{c,i} - \sum_{i \in \mathcal{G}_0} \hat{\omega}_{c,i} \hat{\delta}_{c,i},$$

where $\hat{\delta}_{c,i}$ is given by either the DID, SC, or SDID form above. Control weights for DID are simply $\hat{\omega}_{c,i}^{\text{did}} = N_{\text{co}}^{-1}$ whereas SC and SDID weights are estimated using their respective minimisation problems (Abadie, Diamond and Hainmueller, 2010; Arkhangelsky et al., 2021). Under DiD with a continuous treatment, the counterfactual is common across cohorts: every treated cohort's outcomes must trend in parallel with the donor pool before treatment (Callaway, Goodman-Bacon

and Sant'Anna, 2024*a*). Under SC and SDID, unit weights are recomputed for each cohort separately. This relaxes parallel trends in favour of a latent factor assumption (see Section 3). SDID further departs from SC by estimating time weights $\hat{\lambda}$, which weight pre-treatment periods whose control outcomes most closely resemble the post-treatment level for control units.

A. Potential Outcomes and the Average Treatment Effect on Treated Units

In terms of potential outcomes, the potential outcome for unit i at time t under dose d is denoted by $Y_{it}(d)$. The level treatment effect of changing a unit's dose from the lowest dose in the panel, d_L, to some dose d is

$$Y_{it}(d) - Y_{it}(d_L).$$

$Y_{it}(d)$ is the outcome of unit i in a world in which the unit experiences dose d, whereas $Y_{it}(d_L)$ is the outcome in a world in which the unit instead experiences the lowest possible dose d_L. The inclusion of t allows treatment effects to vary over time, such that potential outcomes define a causal effect for every unit in every time period. The $ATT(d|d)$ is the expectation of this level effect after treatment time T_{pre}, conditional on realised dosage d. The first term is the average outcome of units with realised dosage d after time T_{pre}. The second term $E[Y_{t>T_{\text{pre}}}(d_L)|D = d]$ is a counterfactual. It is the average post-treatment outcome that units who received dose d would have had if they had instead been exposed to the baseline dose d_L,

$$\begin{aligned} ATT(d|d) &= E\big[Y_{t>T_{\text{pre}}}(d) - Y_{t>T_{\text{pre}}}(d_L)|D = d\big] \\ &= E\big[Y_{t>T_{\text{pre}}}(d)|D = d\big] - E\big[Y_{t>T_{\text{pre}}}(d_L)|D = d\big]\,. \end{aligned}$$

B. Causal Response Parameters

This section extends the causal parameter to understand how the average treatment effect on the treated changes if a unit experienced a dose other than its realised dosage. This requires an additional restriction on dose-response heterogeneity which is equivalent to the *Strong Parallel Trends* assumption in the difference-in-differences with a continuous treatment literature (Callaway, Goodman-Bacon and Sant'Anna, 2024*a*).

Assumption 1 (Dose-Response Homogeneity)**.** *To identify* ATT$(d|d')$*, for all* $d \in \mathcal{D}$ *and* $l \in \mathcal{D}$

$$\mathbb{E}\big[Y_t(d) - Y_t(d_L) \;\big|\; D = l\big] = \mathbb{E}\big[Y_t(d) - Y_t(d_L) \;\big|\; D = d\big].$$

Assumption 1 rules out systematic differences in average dose effects across realised-dose groups: the effect at dose d for units with realised dose l equals the effect for units with realised dose d. This assumption extrapolates the average

effect of a particular dose from the treated sub-population to the entire population. Violations of this assumption can arise when a confounder makes it such that assignment of dose is correlated with treatment effects.

Define the $ATT(d|d')$ as the average effect of dose d on units that actually experienced dose d'

$$ATT(d|d') = E[Y_{t>T_{\text{pre}}}(d) - Y_{t>T_{\text{pre}}}(d_L)|D = d']$$

which measures the treatment effect post-treatment of changing from lowest possible dose to dose d for a given unit which actually received treatment dose d'. These definitions can be extended to average causal response parameters, or the marginal effects of increasing treatment dose, such that the average causal response of the treated is,

$$ACRT(d|d') \;=\; \frac{\mathrm{d}\, ATT(l|d')}{\mathrm{d}l}\bigg|_{l=d} \;=\; \frac{\mathrm{d}\,\mathbb{E}\big[Y(l)\,\big|\,D=d'\big]}{\mathrm{d}l}\bigg|_{l=d}$$

To further summarise this measure, take the average of the $ACRT(d|d')$ which is defined as the $ACRT^{\circ}$. The $ACRT^{\circ}$ is comparable to the β parameter commonly observed in $TWFE$, except the estimator in $TWFE$ has been shown to be a biased estimator except under special distributions (Callaway, Goodman-Bacon and Sant'Anna, 2024*a*). The $ACRT^{\circ}$ is defined as:

$$ACRT^{\circ} \;=\; E\big[\, ACRT\big(D|D\big)|D > d_L\big]$$

Figure 2 shows an example of the dose-response using an example curve of the $ATT(d|d)$. The ACRT at a specific dose is the tangent to the $ATT(d|d)$. Interpreting the $ATT(d|d)$ correctly is important. For example, even if all values of the $ATT(d|d)$ are positive, it may still be that the average effects of a change in dosage $ACRT(d|d)$ is negative or zero (Callaway, Goodman-Bacon and Sant'Anna, 2024*a*).

C. Second-Stage Estimation of the Dose-Response

Algorithm 1 estimates treatment effects for each cohort $\{(\hat{\tau}_c, d_c)\}_{c=1}^{C}$, where d_c is the realized dose for cohort c and $\hat{\tau}_c$ is the repeated-cohort SDID estimate computed on the sub-panel $\mathcal{G}_0 \cup \mathcal{G}_c$. For each realized cohort dose d_c, the first-stage target is

$$\tau_c = ATT(d_c|d_c) = \frac{1}{T_{\text{post}}} \sum_{t=T_{\text{pre}}+1}^{T} \mathbb{E}[Y_t(d_c) - Y_t(d_L)|D = d_c]\,.$$

Thus, the first stage provides estimates of the curve $f(d) = ATT(d|d)$, $d \in \mathcal{D}^{+}$, evaluated at the realized cohort doses $\{d_c\}_{c=1}^{C}$. By Assumption 1, treatment

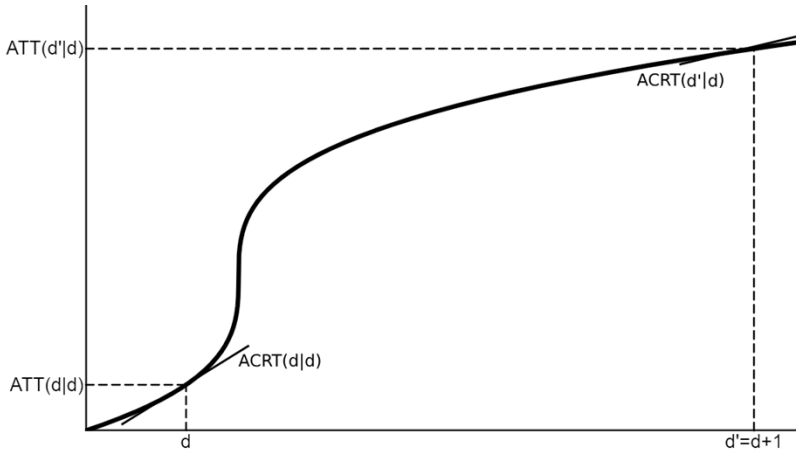


Figure 2. : The height of the graph represents $ATT(\cdot|d)$, whereas the x-axis represents the dose. This shows the dose-response for a unit with a realised dosage d. The curve itself is not based on data. These parameters are described graphically with an example curve for treatment, where the $ACRT(\cdot|d)$ is the slope of the tangent line.

effects are homogeneous across realized-dose groups, such that $ATT(d|d') = ATT(d|d)$ for all $d, d' \in \mathcal{D}^+$. Under this restriction, a curve fitted to the first-stage objects $ATT(d|d)$ is also a curve for the causal object $ATT(d|d')$. Consequently, differentiating the fitted curve with respect to d identifies

$$ACRT(d|d') = \frac{\partial ATT(d|d')}{\partial d},$$

and averaging this derivative over the treated-dose distribution identifies

$$ACRT^{\circ} = \mathbb{E}[ACRT(D|D)|D > d_L]\,.$$

In the following section, I consider two second-stage estimation strategies which target parameters of interest including the $ATT(d|d)$, $ACRT(d|d')$ as well as the $ACRT^{\circ}$.

Nonparametric specification.

As the main specification, I fit a smooth nonparametric curve to the estimated dose-response points $\{(d_c, \hat{\tau}(d_c))\}$. Let $\mathcal{H}^m$ denote a class of sufficiently smooth functions on $\mathcal{D}^+$. Define

$$\hat{f} = \underset{g \in \mathcal{H}^m}{\arg\min} \left\{ \int_{\mathcal{D}^+} \big(\hat{\tau}(d_c) - g(d_c)\big)^2 \, \mathrm{d}\hat{F}_D(d_c) + \mu J(g) \right\}, \tag{4}$$

where $\hat{F}_D$ is the empirical distribution of realized treated doses, $J(g)$ is a smoothness penalty, and μ is a tuning parameter. In practice, (4) can be implemented using a generalized additive model (Wood, 2017). Under Assumption 1, $\hat{f}(d_c)$ estimates both $ATT(d|d)$ and $ATT(d|d')$, while its derivative $\hat{f}'(d_c)$ estimates $ACRT(d|d')$. A sample analogue of the average causal response is then

$$\widehat{ACRT}^{\circ} = \int_{\mathcal{D}^+} \hat{f}'(d_c)\, \mathrm{d}\hat{F}_D(d_c).$$

Linear specification.

Suppose the dose-response is linear: $ATT(d|d') = \alpha + \beta d$. Under Assumption 1, this is equivalent to assuming $ATT(d|d) = \alpha + \beta d$. The second stage is then estimated by the regression

$$(\hat{\alpha}, \hat{\beta}) = \underset{(a,b)\in\mathbb{R}^2}{\arg\min} \sum_{c=1}^{C} \big(\hat{\tau}_c - a - b d_c\big)^2. \tag{5}$$

Since the slope is constant across doses, $ACRT(d|d') = \beta$ for all $d, d' \in \mathcal{D}^+$, and therefore $ACRT^{\circ} = \beta$. In this case, $\hat{\beta}$ directly estimates the average causal response of a one-unit increase in treatment on the treated.

Remark 1 (OLS need not identify $ACRT^{\circ}$ under nonlinearity)**.** *Let $\mu := F_D(\cdot \mid D > d_L)$ denote the treated-dose distribution on $\mathcal{D}^+$. Suppose the probability limit of the OLS slope admits the representation*

$$\beta^*_{ols} = \frac{\mathrm{Cov}(f(D), D \mid D > d_L)}{\mathrm{Var}(D \mid D > d_L)} = \int_{\mathcal{D}^+} f'(d)\, w(d)\, d\mu(d),$$

where, by the Yitzhaki (1996) representation,

$$w(d) = \frac{\big(\mathbb{E}[D \mid D \geq d,\ D > d_L] - \mathbb{E}[D \mid D > d_L]\big)\Pr(D \geq d \mid D > d_L)}{\mathrm{Var}(D \mid D > d_L)}.$$

Also define

$$ACRT^{\circ} = \int_{\mathcal{D}^+} f'(d)\, d\mu(d).$$

Then

$$\beta^*_{ols} - ACRT^{\circ} = \int_{\mathcal{D}^+} f'(d)\big(w(d) - 1\big)\, d\mu(d).$$

Hence nonlinearity of f alone does not imply that $\beta^*_{ols} \neq ACRT^{\circ}$. A sufficient condition for inconsistency is that f' and $f'w$ are μ-integrable and

$$\int_{\mathcal{D}^+} f'(d)\big(w(d) - 1\big)\, d\mu(d) \neq 0.$$

Under this condition,

$$\beta^*_{ols} \neq ACRT^{\circ}.$$

*Thus, under a nonlinear dose-response, OLS generally recovers a weighted average of marginal effects rather than the average causal response. This is the same weighting critique emphasized by Callaway, Goodman-Bacon and Sant'Anna (2024*a*).*

II. Monte Carlo Simulations

A. Data Generating Process

In this section, I conduct a Monte Carlo simulation under a null effect study. The simulations study the finite-sample performance of SDID, SC, and DiD with a continuous treatment. To make this setting more challenging, treatment intensity and unit-specific trends are driven using latent unit-factors calibrated from data. I use the Penn World Tables and the Current Population Survey data to construct a range of simulation designs based on the following data-generating process.

Let Y^*_{it} denote the observed outcome. I fit a rank-four factor model, and decompose it into a fixed effects term F_{it} and an interactive term M_{it}.

$$L := \underset{L:\,\mathrm{rank}(L)=4}{\arg\min} \sum_{it} (Y^*_{it} - L_{it})^2,$$

$$F_{it} = \alpha_i + \beta_t = \frac{1}{T}\sum_{l=1}^{T} L_{il} + \frac{1}{N}\sum_{j=1}^{N} L_{jt} - \frac{1}{NT}\sum_{i,t} L_{it},$$

$$M_{it} = L_{it} - F_{it}.$$

I draw noise as $E_i \sim \mathcal{N}(0, \Sigma)$, where Σ is estimated by fitting an AR(2) model to the residuals $Y^*_{it} - L_{it}$. For treatment assignment, treated units are chosen such that

$$\pi_i = \pi(\alpha_i, \mathbf{M}_i;\, \hat{\phi}) = \frac{\exp(\hat{\phi}_\alpha \alpha_i + \hat{\phi}_M \mathbf{M}_i)}{1 + \exp(\hat{\phi}_\alpha \alpha_i + \hat{\phi}_M \mathbf{M}_i)},$$

$$D_i|\pi_i \;\sim\; \pi_i \times \mathrm{Bernoulli}(\pi_i),$$

where α_i and $\mathbf{M}_i$ are the unit fixed effect and interactive factor component respectively. If a unit is assigned into treatment, then its realised dosage is the latent unit propensity π_i. Beyond this, I also introduce confounding between the latent propensity π_i and unit-specific slopes. Units with a high π_i are: more likely to be treated, more likely to have a high dose assigned, and more likely to have steeper upward trends across time. This slope term is specified as,

$$s_i = \left(\pi_i - \tfrac{1}{2}\right) + \mathbf{1}\{D_i = 0\}\eta_i,$$

$$\eta_i \sim \mathcal{N}(0,\; 4\max\{\mathrm{Var}(\pi_j|D_j > d_L) - \mathrm{Var}(\pi_j|D_j = d_L),\, 0\})\,,$$

where η_i ensures control units receive a data-calibrated mean-zero perturbation. This ensures donor pool slopes are sufficiently dispersed to match the trends of treated units. The full specification of the panel is

$$Y_{it} = F_{it} + M_{it} + s_i t + E_{it}$$

where $ACRT^{\circ} = 0$ is the true dose-response slope. Under DiD with a continuous treatment, violations of parallel trends and actual treatment effects are indistinguishable. A unit with a higher slope relative to the average control is observationally equivalent to one whose outcomes increase as a result of treatment. As a result, DiD with a continuous treatment is predicted to attribute the dose-correlated pre-trends to a treatment effect, producing upward bias in the estimated slope.

Table 1—: Root mean squared error (RMSE) and bias of the dose-response slope estimate. Bias and RMSE are multiplied by 10 for readability. $\bar{N}_1$ is the average number of treated units across replications. Results are based on 500 replications. $T_{post} = 10$ and cDiD refers to DiD with a continuous treatment (Callaway, Goodman-Bacon and Sant'Anna, 2024*a*).

		RMSE				Bias			
Simulator	$\bar{N}_1$	SDID	SC	DiD	cDiD	SDID	SC	DiD	cDiD
Current Population Survey ($N = 50,\ T = 40$)									
Baseline	8.2	**1.68**	3.40	12.31	12.31	0.57	1.63	12.14	12.14
Outcome model									
No corr.	8.2	**1.69**	3.41	12.33	12.33	0.58	1.64	12.15	12.15
No M	8.2	**1.61**	2.77	9.83	9.83	0.73	1.35	9.81	9.81
No F	8.2	**1.67**	2.56	12.31	12.31	0.56	1.10	12.14	12.14
Noise and slope only	8.2	**1.60**	1.81	9.83	9.83	0.71	0.65	9.81	9.81
No noise or slope	8.2	**0.25**	1.14	3.06	3.06	0.18	0.58	2.33	2.33
No slope	8.2	**1.34**	1.72	3.13	3.13	0.52	1.17	2.35	2.35
Assignment process									
Gun law	25.0	**2.33**	4.80	11.95	11.95	2.09	4.64	11.90	11.90
Abortion	17.9	**1.20**	3.32	9.96	9.96	0.76	1.34	9.93	9.93
Outcome variable									
Hours	7.9	7.79	**7.77**	58.51	58.51	4.38	4.72	58.32	58.32
Urate	8.2	**12.24**	13.53	70.09	70.09	5.20	8.29	69.71	69.71
Resample N=200	31.6	**0.74**	0.76	11.50	11.81	0.08	−0.25	11.47	11.77
Resample N=400	64.6	**0.49**	0.60	11.23	11.81	−0.00	−0.32	11.21	11.79
Penn World Tables ($N = 111,\ T = 48$)									
Democracy	28.6	**1.07**	3.56	10.42	10.42	0.67	2.38	10.24	10.24
Education	13.9	**1.39**	7.00	6.63	6.63	0.80	4.89	6.19	6.19

Table 1 reports the root mean squared error (RMSE) and bias from the Monte Carlo simulations. Each row varies the data-generating process (DGP), by omitting structural terms or changing the treatment assignment mechanism. Column headers denote the estimators evaluated, where 'cDiD' refers to DiD with a

continuous treatment from Callaway, Goodman-Bacon and Sant'Anna (2024*a*). This differs from DiD under a repeated-cohort specification, as 'cDiD' does not group units which share the same dosage, which changes the second-stage OLS estimates.

In Table 1, SDID consistently achieves the lowest RMSE and bias across almost all DGPs. Conversely, DiD estimators have substantially higher RMSE and bias relative to SC and SDID. DiD performs particularly poorly when the latent-slope term is included, whereas SDID and SC remain robust to its inclusion. DiD still suffers a bias under the no slope specification, as confoundedness is introduced by treatment assignment being determined by the interactive factor term M_{it}. Finally, both SDID and SC exhibit improved performance under resampling. For SDID, this effectively drives the average bias toward zero. Table A1 in the appendix shows RMSE and bias when dosage is binned into $K = 5$ bins. It slightly increases the RMSE and bias across estimators, except for SC where it marginally reduces the RMSE under the *Gun Law* and *Democracy* designs. Table A2 shows how RMSE and bias changes when the number of total treated units is limited to $\frac{N}{5}$, where *Gun Law* and *Abortion* designs have the most significant reduction in bias.

Table 2 reports empirical coverage rates for the nominal 95 percent coverage rate. The intervals used to determine coverage are nominal 95% intervals from the second-stage OLS regression of $\hat{\tau}_c$ on d_c. These intervals should be interpreted as a finite-sample benchmark rather than as a direct consequence of the formal results. Besides the coverage shown, I evaluated alternative inference procedures including block, jackknife, and placebo bootstrapping on each $\hat{\tau}_d$, and across the original panel (see Algorithms A1-A4 in the appendix). Across all bootstrap specifications, empirical coverage converges to 100%. This is likely due to cross-cohort dependence in finite samples: common shocks within the donor pool are treated as independent sources of variation, resulting in a bootstrap distribution for the second-stage slope that is too dispersed relative to the true sampling distribution.

Using standard errors from OLS, Table 2 demonstrates that SDID attains coverage rates close to the nominal 95% level particularly using the baseline, outcome, and resampling simulation designs. In contrast, DiD estimators have severe under-coverage that is close to zero. SC performs reasonably well, but its coverage is less stable across simulation designs and deteriorates particularly when dose support is continuous. Grouping units into dose bins generally improves inference for both SDID and SC in small samples. The main exception is for SDID under *Resampling*, where the estimator appears to perform better using a continuous dose specification. In the *Gun Law* design, which has a particularly high proportion of treated units, SDID and SC have substantial undercoverage. This shows that the estimator performs poorly when the donor pool is too limited to construct a reliable counterfactual, even if this tradeoff improves the degrees of freedom of the second-stage OLS estimator. Table A3 in the appendix restricts the total number of treated units in simulations to be no more than $\frac{1}{5}$ of the total units per panel N. This improves coverage for designs which had a high number of

Table 2—: Coverage results for nominal 95% coverage rates. Doses are grouped into $K = 5$ bins. $\bar{N}_1$ is the average number of treated units across replications. Results are based on 500 replications. $T_{post} = 10$, and cDiD refers to DiD with a continuous treatment (Callaway, Goodman-Bacon and Sant'Anna, 2024*a*).

Simulator	Dose Grouping	$\bar{N}_1$	SDID	SC	DiD	cDiD
Current Population Survey ($N = 50,\ T = 40$)						
Baseline	Binned	8.2	**0.93**	0.87	0.04	0.02
	Continuous		**0.91**	0.82	0.01	0.01
Assignment process						
Gun law	Binned	25.0	**0.54**	0.29	0.00	0.00
	Continuous		**0.36**	0.00	0.00	0.00
Abortion	Binned	17.9	0.86	**0.91**	0.05	0.00
	Continuous		**0.80**	0.64	0.00	0.00
Outcome variable						
Hours	Binned	7.9	**0.96**	0.92	0.00	0.00
	Continuous		**0.90**	0.87	0.00	0.00
Urate	Binned	8.2	**0.94**	0.89	0.03	0.02
	Continuous		**0.92**	0.80	0.01	0.01
Resample, $N = 200$	Binned	31.6	**0.92**	0.92	0.00	0.00
	Continuous		**0.92**	0.90	0.00	0.00
Resample, $N = 400$	Binned	64.6	**0.94**	0.90	0.00	0.00
	Continuous		**0.96**	0.92	0.00	0.00
Penn World Tables ($N = 111,\ T = 48$)						
Democracy	Binned	28.6	0.88	**0.92**	0.26	0.04
	Continuous		**0.83**	0.56	0.02	0.02
Education	Binned	13.9	**0.92**	0.84	0.56	0.39
	Continuous		**0.89**	0.42	0.33	0.33

treated units, though the empirical coverage still sits below 95%. Further, Table A4 in the appendix compares simulations with and without the slope term, and shows that coverage for DiD improves when this slope term is removed.

Overall, these simulations show that SDID with a continuous treatment attains impressive coverage and bias even under challenging simulation designs which have a limited number of observations and high confounding between dosage, treatment assignment, and unit-specific slopes.

III. Formal Results

This section implements the formal results and regularity conditions from Arkhangelsky et al. (2021), extending the assumptions to SDID with a continuous treatment. Assume the data generating process (DGP) of the panel follows a latent factor model where γ_i is a vector of latent unit factors of dimension R, υ_t

is a vector of latent time factors of dimension R, and ε_{it} is the error term.

$$Y_{it} = \gamma_i \upsilon_t^\top + \tau_{it}(D_i) W_{it} + \varepsilon_{it}.$$

τ_{it} is a function of dose D_i, and can vary across time. The treatment matrix is specified as $\boldsymbol{W}_{it} = \mathbf{1}\{D_i > d_L, t > T_{\text{pre}}\}$ such that treatment effects τ_{it} only affect treated outcomes after time T_{pre}. This can be rewritten in matrix notation

$$\boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{\tau} \circ \boldsymbol{W} + \boldsymbol{E}, \qquad \text{where } \boldsymbol{L} = \boldsymbol{\Gamma}\boldsymbol{\Upsilon}^\top.$$

such that $\boldsymbol{L}$, the systematic component, is characterised by a factor model with latent unit factors $\boldsymbol{\Gamma}$ and latent time factors $\boldsymbol{\Upsilon}$. $\boldsymbol{L}$ can be further defined via a singular value decomposition $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$ such that $\boldsymbol{\Gamma} = \boldsymbol{U}\boldsymbol{D}^{1/2}$ and $\boldsymbol{\Upsilon}^\top = \boldsymbol{D}^{1/2}\boldsymbol{V}^\top$. Let $d_L = \min_i(D_i)$, and let

$$d_L < d_1 < d_2 < \cdots < d_C$$

denote the distinct realized treated doses strictly above d_L. Dose cohorts are indexed by $\mathbf{C} := \{1, \ldots, C\}$, where cohort c is defined by the common realized dose d_c,

$$\mathcal{G}_0 = \{i : D_i = d_L\}, \qquad \mathcal{G}_c = \{i : D_i = d_c\}, \quad c \in \mathbf{C}.$$

Because each d_c is a realized treated dose, every realized cohort is nonempty, so

$$|\mathcal{G}_c| > 0 \qquad \text{for all } c \in \mathbf{C}.$$

Let the number of control units be denoted by $N_{co} := |\mathcal{G}_0|$. The target parameter is specified as

$$\tau_c = \frac{1}{|\mathcal{G}_c|\, T_{\text{post}}} \sum_{i \in \mathcal{G}_c} \sum_{t=T_{\text{pre}}+1}^{T} \tau_{it}.$$

For each $c \in \mathbf{C}$, let $\mathcal{I}_c := \mathcal{G}_c$ denote the set of treated unit indices belonging to dose cohort c. Let subscript c denote cohort-specific parameters, weights, or estimators; whereas superscript (c) denotes a restriction to units in a treated cohort. More formally, for some vector x or matrix X: if $x = (x_i)_{i=1}^N \in \mathbb{R}^N$ is a unit-indexed vector, then superscript (c) denotes the following

$$x^{(c)} := (x_i)_{i \in \mathcal{I}_c} \in \mathbb{R}^{|\mathcal{I}_c|};$$

whereas if $X = (X_{it})_{i=1,\ldots,N;\, t=1,\ldots,T} \in \mathbb{R}^{N \times T}$ is a matrix with unit and time dimensions, then

$$X^{(c)} := (X_{it})_{i \in \mathcal{I}_c,\, t=1,\ldots,T} \in \mathbb{R}^{|\mathcal{I}_c| \times T}.$$

A. Definition of Oracle and Control Weights

In this section, I define and use oracle weights to provide a decomposition of the estimation error of the estimator for each sub-panel. Assume the error matrix $\boldsymbol{E_i}$ has homoskedastic errors across units $i = 1, \ldots, N$ such that $Var[\boldsymbol{E_i}] = \Sigma \in \mathbb{R}^{T \times T}$. Partition the $\boldsymbol{Y}$ and $\boldsymbol{\Sigma}$ matrices such that:

$$Y^{(c)} = \begin{pmatrix} Y_{\text{co,pre}} & Y_{\text{co,post}} \\ Y^{(c)}_{\text{tr,pre}} & Y^{(c)}_{\text{tr,post}} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{\text{pre,pre}} & \Sigma_{\text{pre,post}} \\ \Sigma_{\text{post,pre}} & \Sigma_{\text{post,post}} \end{pmatrix}.$$

where $Y_{\text{co,pre}}$ is $N_{\text{co}} \times T_{\text{pre}}$, $Y_{\text{co,post}}$ is $N_{\text{co}} \times T_{\text{post}}$, $Y^{(c)}_{\text{tr,pre}}$ is $|\mathcal{G}_c| \times T_{\text{pre}}$, and $Y^{(c)}_{\text{tr,post}}$ is $|\mathcal{G}_c| \times T_{\text{post}}$. The same partitioning notation for Y can also be extended to $L^{(c)}$, $W^{(c)}$, τ, and $E^{(c)}$. For each cohort $c \in \mathbf{C}$, let $\omega_c \in \mathbb{R}^{N_{\text{co}}+|\mathcal{G}_c|}$ denote the unit-weight vector on the sub-panel $\mathcal{P}_c = \mathcal{G}_0 \cup \mathcal{G}_c$, and let $\lambda \in \mathbb{R}^T$ denote the time-weight vector which is the same across panels. Partition ω_c and λ conformably as

$$\omega_c = \begin{pmatrix} \omega_{\text{co},c} \\ \omega_{\text{tr},c} \end{pmatrix}, \quad \lambda = \begin{pmatrix} \lambda_{\text{pre}} \\ \lambda_{\text{post}} \end{pmatrix},$$

where $\omega_{\text{co},c} \in \mathbb{R}^{N_{\text{co}}}$, $\omega_{\text{tr},c} \in \mathbb{R}^{|\mathcal{G}_c|}$, $\lambda_{\text{pre}} \in \mathbb{R}^{T_{\text{pre}}}$, and $\lambda_{\text{post}} \in \mathbb{R}^{T_{\text{post}}}$ for each cohort $c \in \mathbf{C}$. For any weights $\omega_c \in \Omega_c \subset \mathbb{R}^{N_{\text{co}}+|\mathcal{G}_c|}$ and $\lambda \in \Lambda$, define the weighted double-differences for each cohort as:

$$\hat{\tau}_c(\omega_c, \lambda) = \omega^\top_{\text{tr},c} Y^{(c)}_{\text{tr,post}} \lambda_{\text{post}} - \omega^\top_{\text{co},c} Y_{\text{co,post}} \lambda_{\text{post}} - \omega^\top_{\text{tr},c} Y^{(c)}_{\text{tr,pre}} \lambda_{\text{pre}} + \omega^\top_{\text{co},c} Y_{\text{co,pre}} \lambda_{\text{pre}}.$$

For any $\omega_c \in \Omega_c$ and $\lambda \in \Lambda$, suppose the treated-unit weights and post-treatment time weights are uniform, so that

$$\omega_{\text{tr},c} = |\mathcal{G}_c|^{-1} \mathbf{1}_{|\mathcal{G}_c|} \qquad \text{and} \qquad \lambda_{\text{post}} = T^{-1}_{\text{post}} \mathbf{1}_{T_{\text{post}}}.$$

then by definition of τ_c,

$$\omega^\top_{\text{tr},c} \tau^{(c)}_{\text{tr,post}} \lambda_{\text{post}} = \tau_c.$$

where $\tau^{(c)}_{\text{tr,post}}$ denotes the array of treatment effects τ_{it} restricted to post-treatment, for treated units in cohort c. The error of any weighted double-differencing estimator can then be established by substituting in the Y matrix,

$$\hat{\tau}_c(\omega_c, \lambda) - \tau_c = \underbrace{\omega^\top_{\text{tr},c} L^{(c)}_{\text{tr,post}} \lambda_{\text{post}} - \omega^\top_{\text{co},c} L_{\text{co,post}} \lambda_{\text{post}} - \omega^\top_{\text{tr},c} L^{(c)}_{\text{tr,pre}} \lambda_{\text{pre}} + \omega^\top_{\text{co},c} L_{\text{co,pre}} \lambda_{\text{pre}}}_{\text{bias } B(\omega_c, \lambda)}$$
$$+ \underbrace{\omega^\top_{\text{tr},c} E^{(c)}_{\text{tr,post}} \lambda_{\text{post}} - \omega^\top_{\text{co},c} E_{\text{co,post}} \lambda_{\text{post}} - \omega^\top_{\text{tr},c} E^{(c)}_{\text{tr,pre}} \lambda_{\text{pre}} + \omega^\top_{\text{co},c} E_{\text{co,pre}} \lambda_{\text{pre}}}_{\text{noise } \varepsilon(\omega_c, \lambda)}.$$

Define oracle weights $\tilde{\omega}_c$ and $\tilde{\lambda}$ as the expectation of the objective functions (1) and (3) by setting

$$(\tilde{\omega}_{0,c}, \tilde{\omega}_c) = \underset{\omega_{0,c}\in\mathbb{R},\, \omega_c\in\Omega_c}{\arg\min}\ \mathbb{E}[\ell_{\text{unit}}(\omega_{0,c}, \omega_c)], \qquad (\tilde{\lambda}_0, \tilde{\lambda}) = \underset{\lambda_0\in\mathbb{R},\, \lambda\in\Lambda}{\arg\min}\ \mathbb{E}[\ell_{\text{time}}(\lambda_0, \lambda)].$$

Unit weights depend on pre-treatment periods only, and time weights depend only on control units, such that the objective functions can be expressed using only the systematic component $\boldsymbol{L}$ and the error term $\boldsymbol{E}$. Substituting in $\boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{E}$, and taking the expectation of the objective function results in the simplified characterisation:

$$(\tilde{\omega}_{0,c}, \tilde{\omega}_c) = \underset{\omega_{0,c}\in\mathbb{R},\, \omega_c\in\Omega_c}{\arg\min} \left\|\omega_{0,c} + \omega_{\text{co},c}^\top L_{\text{co,pre}} - \omega_{\text{tr},c}^\top L_{\text{tr,pre}}^{(c)}\right\|_2^2 + \left(\text{tr}(\Sigma_{\text{pre,pre}}) + \zeta^2 T_{\text{pre}}\right) \|\omega_c\|_2^2,$$

$$(\tilde{\lambda}_0, \tilde{\lambda}) = \underset{\lambda_0\in\mathbb{R},\, \lambda\in\Lambda}{\arg\min} \|\lambda_0 + L_{\text{co,pre}}\lambda_{\text{pre}} - L_{\text{co,post}}\lambda_{\text{post}}\|_2^2 + \|\tilde{\Sigma}\lambda\|_2^2.$$

where

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma_{\text{pre,pre}} & -\Sigma_{\text{pre,post}} \\ -\Sigma_{\text{post,pre}} & \Sigma_{\text{post,post}} \end{pmatrix}.$$

Note the error structure does not depend on cohorts given that the rows of the error matrix are assumed to be independently and identically distributed with the common covariance matrix Σ. The total estimation error $\hat{\tau}_c^{\text{sdid}} - \tau_c$ is decomposed by adding and subtracting the infeasible oracle estimator $\hat{\tau}_c(\tilde{\omega}_c, \tilde{\lambda})$, which yields

$$\hat{\tau}_c^{\text{sdid}} - \tau_c = \underbrace{\varepsilon(\tilde{\omega}_c, \tilde{\lambda})}_{\text{oracle noise}} + \underbrace{B(\tilde{\omega}_c, \tilde{\lambda})}_{\text{oracle confounding bias}} + \underbrace{\hat{\tau}_c(\hat{\omega}_c, \hat{\lambda}) - \hat{\tau}_c(\tilde{\omega}_c, \tilde{\lambda})}_{\text{deviation from oracle}}.$$

This expression shows that oracle noise, confounding bias, and deviation from oracle weights vary across cohorts, as these terms depend on cohort-specific unit weights $\hat{\omega}_c$. In the next section, I specify the assumptions required for each sub-panel to be asymptotically consistent. When Σ is not diagonal then the regularization term shrinks $\tilde{\lambda}$ towards the autoregressive vector

$$\psi = \underset{v\in\mathbb{R}^{T_{\text{pre}}}}{\arg\min} \left\|\tilde{\Sigma}\begin{pmatrix} v \\ \lambda_{\text{post}} \end{pmatrix}\right\|_2 = \Sigma_{\text{pre,pre}}^{-1}\Sigma_{\text{pre,post}}\lambda_{\text{post}}.$$

B. Asymptotic Properties of SDID

This section studies the first-stage large-sample behavior of the cohort-specific SDID estimators on sub-panels $\mathcal{P}_c = \mathcal{G}_0 \cup \mathcal{G}_c$. The cohort-specific estimation error admits the same bias, noise, and deviation decomposition as in the usual SDID problem, thus the role of Assumptions 2–5 is to verify when the regularity

conditions required by Arkhangelsky et al. (2021) hold on each fixed sub-panel. Accordingly, the consistency, asymptotic normality, and pointwise interval-validity statements below should be read as cohort-by-cohort consequences of the corresponding SDID asymptotic theory applied to $\mathcal{P}_c$, rather than as a separate proof from first principles. I define

$$N_{tr} := \sum_{c=1}^{C} |\mathcal{G}_c|$$

as the total number of treated units in the full panel.

Assumption 2 (Properties of Errors)**.** *The rows E_i of the noise matrix are independent and identically distributed Gaussian vectors and the eigenvalues of its covariance matrix Σ are bounded and bounded away from zero.*

Assumption 2 is inherited by every cohort-specific sub-panel $\mathcal{P}_c = \mathcal{G}_0 \cup \mathcal{G}_c$, as each sub-panel is obtained by restricting the full error matrix to a subset of rows. Since the row vectors E_i are assumed to be independent and identically distributed Gaussian vectors with common covariance matrix Σ, the restricted collection $\{E_i : i \in \mathcal{P}_c\}$ remains independent and identically distributed with the same covariance matrix. Consequently, the covariance blocks and the eigenvalue bounds of Σ are invariant across cohorts.

Assumption 3 (Sample Sizes)**.** *I consider a sequence of populations where*

(i) for all c in $\mathbf{C}$ the product $|\mathcal{G}_c|T_{\text{post}}$ goes to infinity, and both N_{co} and T_{pre} go to infinity,

(ii) the ratio $T_{\text{pre}}/N_{\text{co}}$ is bounded and bounded away from zero,

(iii)

$$\frac{N_{\text{co}}}{N_{\text{tr}} T_{\text{post}} \max(N_{\text{tr}}, T_{\text{post}}) \log^2(N_{\text{co}})} \to \infty.$$

Assumption 3(i) is stronger than the usual assumption required for SDID, as it requires that the product between the number of post-treatment periods and the number of treated units for every sub-panel goes to infinity. Assumption 3(ii) holds across sub-panels given that T_{pre} and N_{co} do not vary by dose-cohort. Finally, Assumption 3(iii) holds across every sub-panel given that $|\mathcal{G}_c| \leq N_{tr}$ for all c, where the map $x \mapsto x \max(x, T_{\text{post}})$ is weakly increasing. Every sub-panel satisfies Assumption 3(iii), as the cohort-specific denominator is weakly smaller than the full-panel denominator such that,

$$\frac{N_{\text{co}}}{|\mathcal{G}_c|\, T_{\text{post}} \max\big(|\mathcal{G}_c|, T_{\text{post}}\big) \log^2(N_{\text{co}})} \;\geq\; \frac{N_{\text{co}}}{N_{\text{tr}}\, T_{\text{post}} \max\big(N_{\text{tr}}, T_{\text{post}}\big) \log^2(N_{\text{co}})}.$$

Assumption 4 (Properties of L)**.** *Letting $\sigma_1(\Gamma), \sigma_2(\Gamma), \ldots$ denote the singular values of the matrix Γ in decreasing order and R the largest integer less than $\sqrt{\min(T_{\text{pre}}, N_{\text{co}})}$,*

$$\frac{\sigma_R(L_{\text{co,pre}})}{R} = o\Big(\min\Big\{N_{\text{tr}}^{-1/2}\log^{-1/2}(N_{\text{co}}),\ T_{\text{post}}^{-1/2}\log^{-1/2}(T_{\text{pre}})\Big\}\Big).$$

Assumption 4 is inherited by every sub-panel because it is imposed on the singular values of the common control pre-treatment block $L_{\text{co,pre}}$, which does not vary across cohorts. To compare the rate bounds, note that each realized cohort is nonempty, so $|\mathcal{G}_c| > 0$, and by construction $|\mathcal{G}_c| \le N_{\text{tr}}$ for all $c \in \mathbf{C}$. Since the map $x \mapsto x^{-1/2}$ is decreasing on $(0, \infty)$,

$$|\mathcal{G}_c|^{-1/2} \ge N_{\text{tr}}^{-1/2}.$$

Therefore the cohort-specific bound is weakly larger:

$$\begin{aligned}&\min\Big\{|\mathcal{G}_c|^{-1/2}\log^{-1/2}(N_{\text{co}}),\ T_{\text{post}}^{-1/2}\log^{-1/2}(T_{\text{pre}})\Big\}\\ \ge\ &\min\Big\{N_{\text{tr}}^{-1/2}\log^{-1/2}(N_{\text{co}}),\ T_{\text{post}}^{-1/2}\log^{-1/2}(T_{\text{pre}})\Big\}.\end{aligned}$$

Hence the full-panel rate condition implies the cohort-specific one, and Assumption 4 holds on every sub-panel:

$$\frac{\sigma_R(L_{\text{co,pre}})}{R} = o\Big(\min\Big\{|\mathcal{G}_c|^{-1/2}\log^{-1/2}(N_{\text{co}}),\ T_{\text{post}}^{-1/2}\log^{-1/2}(T_{\text{pre}})\Big\}\Big) \qquad \text{for every } c \in \mathbf{C}.$$

Assumption 5 (Properties of Weights and L)**.** *The oracle unit weights $\tilde{\omega}_c$ for each panel $c \in \mathbf{C}$ satisfy*

$$\|\tilde{\omega}_{\text{co},c}\|_2 = o\Big(\big[(|\mathcal{G}_c|\,T_{\text{post}})\log(N_{\text{co}})\big]^{-1/2}\Big),$$

and

$$\Big\|\tilde{\omega}_{0,c} + \tilde{\omega}_{\text{co},c}^{\top}L_{\text{co,pre}} - \tilde{\omega}_{\text{tr},c}^{\top}L_{\text{tr,pre}}^{(c)}\Big\|_2 =$$
$$o\Big(N_{\text{co}}^{1/4}\big(|\mathcal{G}_c|\,T_{\text{post}}\max(N_{\text{co}}, T_{\text{post}})\big)^{-1/4}\log^{-1/2}(N_{\text{co}})\Big),$$

The oracle time weights of the original panel $\tilde{\lambda}$ satisfy

$$\begin{aligned}\|\tilde{\lambda}_{\text{pre}} - \psi\|_2 &= o\Big(\big[(N_{\text{tr}}T_{\text{post}})\log(N_{\text{co}})\big]^{-1/2}\Big),\\ \Big\|\tilde{\lambda}_0 + L_{\text{co,pre}}\tilde{\lambda}_{\text{pre}} - L_{\text{co,post}}\tilde{\lambda}_{\text{post}}\Big\|_2 &= o\Big(N_{\text{co}}^{1/4}(N_{\text{tr}}T_{\text{post}})^{-1/8}\Big),\end{aligned}$$

and for each $c \in \mathbf{C}$ *the oracle weights jointly satisfy*

$$\tilde{\omega}_{\mathrm{tr},c}^{\top} L_{\mathrm{tr,post}}^{(c)} \tilde{\lambda}_{\mathrm{post}} - \tilde{\omega}_{\mathrm{co},c}^{\top} L_{\mathrm{co,post}} \tilde{\lambda}_{\mathrm{post}} - \tilde{\omega}_{\mathrm{tr},c}^{\top} L_{\mathrm{tr,pre}}^{(c)} \tilde{\lambda}_{\mathrm{pre}} + \tilde{\omega}_{\mathrm{co},c}^{\top} L_{\mathrm{co,pre}} \tilde{\lambda}_{\mathrm{pre}}$$
$$= o\Big((|\mathcal{G}_c|\, T_{\mathrm{post}})^{-1/2} \Big).$$

Assumption 5 is the strongest set of assumptions, as it requires each sub-panel to have treated units and post-treatment periods whose outcomes are similar to control units and pre-treatment periods respectively.

Proposition 1 (Imported first-stage limit theory on each sub-panel)**.** *Fix a cohort* $c \in \mathbf{C}$*. Suppose the cohort-specific sub-panel* $\mathcal{P}_c = \mathcal{G}_0 \cup \mathcal{G}_c$ *satisfies the assumptions required for the asymptotic theory in Arkhangelsky et al. (2021). Then the cohort-specific SDID estimator*

$$\hat{\tau}_c^{\mathrm{sdid}} := \hat{\tau}_c(\hat{\omega}_c, \hat{\lambda})$$

inherits the corresponding pointwise first-stage conclusions on $\mathcal{P}_c$*. In particular,*

$$\hat{\tau}_c^{\mathrm{sdid}} - \tau_c = o_p(1),$$

and, for some cohort-specific asymptotic variance $V_c \in (0, \infty)$*,*

$$\sqrt{|\mathcal{G}_c|\, T_{\mathrm{post}}}\big(\hat{\tau}_c^{\mathrm{sdid}} - \tau_c\big) \xrightarrow{d} \mathcal{N}(0, V_c).$$

If $\hat{V}_c$ *is a consistent variance estimator constructed on* $\mathcal{P}_c$*, then*

$$\frac{\hat{\tau}_c^{\mathrm{sdid}} - \tau_c}{\hat{V}_c^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

and the associated Wald interval

$$\mathrm{CI}_{c,1-\alpha} = \Big[\hat{\tau}_c^{\mathrm{sdid}} - z_{1-\alpha/2} \hat{V}_c^{1/2},\ \hat{\tau}_c^{\mathrm{sdid}} + z_{1-\alpha/2} \hat{V}_c^{1/2} \Big]$$

satisfies

$$\Pr(\tau_c \in \mathrm{CI}_{c,1-\alpha}) \to 1 - \alpha.$$

Likewise, placebo, jackknife, and bootstrap procedures applied separately on $\mathcal{P}_c$ *yield pointwise-valid first-stage inference for* τ_c*, provided their own consistency conditions hold on that same sub-panel.*

Theorem 1 (Joint consistency for a fixed number of cohorts)**.** *If* C *is fixed and*

$$\hat{\tau}_c^{\mathrm{sdid}} - \tau_c = o_p(1) \qquad \textit{for every } c \in \mathbf{C},$$

then

$$\max_{1\le c\le C}\left|\hat{\tau}_c^{\text{sdid}}-\tau_c\right|=o_p(1).$$

Proof. Fix $\varepsilon>0$. For each $c\in\mathbf{C}$,

$$\Pr\Big(\left|\hat{\tau}_c^{\text{sdid}}-\tau_c\right|>\varepsilon\Big)\to 0.$$

Since C is fixed,

$$\Pr\left(\max_{1\le c\le C}\left|\hat{\tau}_c^{\text{sdid}}-\tau_c\right|>\varepsilon\right)\le\sum_{c=1}^{C}\Pr\Big(\left|\hat{\tau}_c^{\text{sdid}}-\tau_c\right|>\varepsilon\Big)\to 0.$$

Hence

$$\max_{1\le c\le C}\left|\hat{\tau}_c^{\text{sdid}}-\tau_c\right|=o_p(1).$$

□

C. Second-Stage Estimation

The first-stage SDID procedure delivers the cohort-level sample $\{(d_c,\hat{\tau}_c)\}_{c=1}^{C}$, where $\hat{\tau}_c$ estimates the diagonal cohort-specific target $\tau_c=ATT(d_c\mid d_c)$.
The second stage treats these cohort-level estimates as noisy evaluations of a continuous dose-response curve. Appendix A provides the conditions under which the realized cohort doses become sufficiently dense in the support of treatment, and the first-stage SDID estimates converge uniformly to their cohort-specific targets. Under these conditions, the second stage recovers both the dose-response curve and average slope summaries. Define the population-level diagonal dose-response curve by

$$f(d):=\frac{1}{T_{\text{post}}}\sum_{t=T_{\text{pre}}+1}^{T}\mathbb{E}[Y_t(d)-Y_t(d_L)\mid D=d],\qquad d\in\mathcal{D}^{+},$$

and let

$$S^{\circ}:=\int_{\mathcal{D}^{+}}f'(d)\,dH_D(d),$$

where H_D denotes the limiting design distribution of the realized cohort doses. If $H_D=F_D(\cdot\mid D>d_L)$, then S° coincides with the treated-unit average causal response

$$ACRT^{\circ}=\mathbb{E}\big[f'(D)\mid D>d_L\big].$$

Assumption 6 (Second-stage regularity)**.** *The diagonal dose-response curve f is continuously differentiable on $\mathcal{D}^{+}$. The realized cohort doses $\{d_c\}_{c=1}^{C}$ become dense in $\mathcal{D}^{+}$, and their empirical distribution converges weakly to a probability*

distribution H_D supported on $\mathcal{D}^+$. Moreover,

$$\max_{1\le c\le C} |\hat{\tau}_c - f(d_c)| = o_p(1).$$

Assumption 6 collects the key regularity conditions needed for second-stage estimation. The appendix shows this follows from the first-stage SDID asymptotic theory, together with support conditions on the dose-cohorts.

Linear specification

Suppose the diagonal dose-response is exactly linear,

$$f(d) = \alpha + \beta d, \qquad d \in \mathcal{D}^+.$$

Then the second-stage estimator is the OLS regression

$$(\hat{\alpha}, \hat{\beta}) = \underset{(a,b)\in\mathbb{R}^2}{\arg\min} \sum_{c=1}^{C} (\hat{\tau}_c - a - b d_c)^2.$$

Theorem 2 (Consistency of the linear second stage)**.** *Suppose Assumption 6 holds and $f(d) = \alpha + \beta d$ for all $d \in \mathcal{D}^+$. Then*

$$\hat{\alpha} \xrightarrow{p} \alpha \qquad \textit{and} \qquad \hat{\beta} \xrightarrow{p} \beta = S^{\circ}.$$

If, in addition, $H_D = F_D(\cdot \mid D > d_L)$, then

$$\hat{\beta} \xrightarrow{p} ACRT^{\circ}.$$

This result shows that under exact linearity the coefficient on dose consistently estimates a constant average causal response. The cohort-specific first-stage estimates share a common donor pool such that conventional second-stage OLS standard errors are not necessarily valid without stronger conditions. Sufficient conditions for heteroskedasticity-robust Wald inference are given in the appendix.

Nonparametric specification

To allow for nonlinear dose-response effects, define the penalized estimator

$$\hat{f} = \underset{g\in\mathcal{H}^m}{\arg\min} \left\{ \frac{1}{C} \sum_{c=1}^{C} \big(\hat{\tau}_c - g(d_c)\big)^2 + \mu_C J(g) \right\}, \tag{6}$$

where $\mathcal{H}^m$ is a class of sufficiently smooth functions, $J(g)$ is a smoothness penalty, and μ_C is a tuning parameter. Let $\hat{S}^{\circ}$ denote the corresponding design-based

average slope estimator,

$$\hat{S}^{\circ} := \frac{1}{C}\sum_{c=1}^{C} \hat{f}'(d_c).$$

Proposition 2 (Consequences of cohort support)**.** *Suppose Assumption 7 holds. Then, for every bounded continuous function $g : \mathcal{D}^+ \to \mathbb{R}$,*

$$\frac{1}{C}\sum_{c=1}^{C} g(d_c) \to \int_{\mathcal{D}^+} g(d)\, dH_D(d).$$

In particular,

$$\frac{1}{C}\sum_{c=1}^{C} f(d_c) \to \int_{\mathcal{D}^+} f(d)\, dH_D(d), \qquad \frac{1}{C}\sum_{c=1}^{C} f'(d_c) \to \int_{\mathcal{D}^+} f'(d)\, dH_D(d).$$

IV. Conclusion

This paper proposes a new estimator which allows researchers to estimate causal effects for continuous treatment without assuming outcomes moved in parallel before treatment. The Monte Carlo study shows SDID with a continuous treatment has a low bias relative to DiD estimators, and provides impressive coverage across a broad range of simulation designs. Inference for the estimator remains sensitive to limited donor pools, and will need to be studied further to understand how these . Revised versions of this working paper will include an empirical application, an extension to event studies similar to Callaway, Goodman-Bacon and Sant'Anna (2024*b*), as well as simulation designs with nonlinear dose specifications, alternative treatment assignment mechanisms and increased sample sizes.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association*, 105(490): 493–505.

Angrist, Joshua D., and Guido W. Imbens. 1995. "Two-stage least squares estimation of average causal effects in models with variable treatment intensity." *Journal of the American Statistical Association*, 90(430): 431–442.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2021. "Synthetic difference-in-differences." *American Economic Review*, 111(12): 4088–4118.

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

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2024*a*. "Difference-in-Differences with a Continuous Treatment." National Bureau of Economic Research 32117.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2024*b*. "Event Studies with a Continuous Treatment." *AEA Papers and Proceedings*, 114: 601–605.

de Chaisemartin, Clément, and Xavier d'Haultfœuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review*, 110(9): 2964–2996.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R.* . 2nd ed., Chapman and Hall/CRC.

Yitzhaki, Shlomo. 1996. "On using linear regressions in welfare economics." *Journal of Business & Economic Statistics*, 14(4): 478–486.

Table A1—: Root mean squared error and bias of the dose-response slope estimate under a binned-dose specification ($K = 5$ bins). $\bar{N}_1$ is the average number of treated units across replications. True slope $\gamma = 0$, $T_{post} = 10$. Bias and RMSE are multiplied by 10 for readability. There are 500 replications and cDiD refers to DiD with a continuous treatment (Callaway, Goodman-Bacon and Sant'Anna, 2024*a*).

		RMSE				Bias			
Simulator	$\bar{N}_1$	SDID	SC	DiD	cDiD	SDID	SC	DiD	cDiD
Current Population Survey ($N = 50, T = 40$)									
Baseline	8.2	**1.70**	3.41	12.49	12.41	0.61	1.52	12.33	12.24
Outcome model									
No corr.	8.2	**1.72**	3.42	12.51	12.43	0.61	1.53	12.35	12.26
No M	8.2	**1.64**	2.85	9.84	9.83	0.78	1.36	9.82	9.81
No F	8.2	**1.70**	2.64	12.49	12.41	0.59	1.14	12.33	12.24
Noise and slope only	8.2	**1.63**	1.88	9.84	9.83	0.76	0.66	9.82	9.81
No noise or slope	8.2	**0.22**	1.07	3.15	3.10	0.17	0.60	2.51	2.43
No slope	8.2	**1.40**	1.62	3.23	3.18	0.58	1.08	2.54	2.45
Assignment process									
Gun law	25.0	**2.49**	3.94	12.14	12.12	2.25	3.66	12.08	12.07
Abortion	17.9	**1.28**	3.33	10.05	10.06	0.88	1.31	10.01	10.02
Outcome variable									
Hours	7.9	8.00	**7.75**	58.44	58.43	4.48	4.43	58.21	58.22
Urate	8.2	**12.50**	12.81	69.80	69.56	5.44	7.04	69.37	69.14
Resample N=200	31.6	**0.71**	0.80	11.94	11.92	0.08	−0.27	11.90	11.87
Resample N=400	64.6	**0.48**	0.60	11.93	11.91	−0.06	−0.29	11.90	11.89
Penn World Tables ($N = 111, T = 48$)									
Democracy	28.6	**1.04**	2.98	10.12	10.18	0.66	1.43	9.90	9.96
Education	13.9	**1.46**	6.90	6.47	6.50	0.82	4.32	6.09	6.12

Table A2—: Root mean squared error (RMSE) and bias of the dose-response slope estimate where the number of treated units is restricted to at most $\frac{N}{5}$ units. $\bar{N}_1$ is the average number of treated units across simulations. Bias and RMSE are multiplied by 10 for readability. Results are based on 500 replications. $T_{post} = 10$ and cDiD refers to DiD with a continuous treatment (Callaway, Goodman-Bacon and Sant'Anna, 2024*a*).

		RMSE				Bias			
Simulator	$\bar{N}_1$	SDID	SC	DiD	cDiD	SDID	SC	DiD	cDiD
Current Population Survey ($N = 50,\ T = 40$)									
Baseline	7.9	**1.67**	3.36	12.32	12.32	0.54	1.54	12.14	12.14
Outcome model									
No corr.	7.9	**1.69**	3.37	12.34	12.34	0.54	1.55	12.16	12.16
No M	7.9	**1.61**	2.71	9.82	9.82	0.72	1.26	9.80	9.80
No F	7.9	**1.67**	2.54	12.32	12.32	0.52	1.06	12.14	12.14
Noise and slope only	7.9	**1.60**	1.80	9.82	9.82	0.70	0.64	9.80	9.80
No noise or slope	7.9	**0.25**	1.12	3.09	3.09	0.18	0.56	2.34	2.34
No slope	7.9	**1.35**	1.70	3.17	3.17	0.51	1.13	2.35	2.35
Assignment process									
Gun law	10.0	**2.21**	3.89	12.21	12.21	1.62	3.60	12.02	12.02
Abortion	10.0	**1.36**	2.33	10.04	10.04	0.88	0.91	9.94	9.94
Outcome variable									
Hours	7.8	**7.90**	7.97	58.52	58.52	4.31	4.77	58.32	58.32
Urate	7.9	**12.35**	13.40	70.20	70.20	5.16	8.08	69.81	69.81
Resample N=200	31.5	**0.74**	0.77	11.50	11.81	0.08	−0.24	11.47	11.77
Resample N=400	64.6	**0.48**	0.59	11.23	11.81	−0.01	−0.33	11.21	11.79
Penn World Tables ($N = 111,\ T = 48$)									
Democracy	22.0	**1.19**	3.17	10.49	10.49	0.64	2.14	10.21	10.21
Education	13.9	**1.39**	7.00	6.63	6.63	0.80	4.89	6.19	6.19

Table A3—: Coverage results for nominal 95% coverage rates. $\bar{N}_1$ is the average number of treated units across replications, restricted to at most $\frac{N}{5}$ units. Bias and RMSE are multiplied by 10 for readability. Results are based on 500 replications. $T_{post} = 10$ and cDiD refers to DiD with a continuous treatment.

Simulator	Dose Grouping	$\bar{N}_1$	SDID	SC	DiD	cDiD
Current Population Survey ($N = 50,\ T = 40$)						
Baseline	Binned	7.9	**0.95**	0.88	0.04	0.02
	Continuous		**0.92**	0.84	0.01	0.01
Assignment process						
Gun law	Binned	10.0	**0.84**	0.68	0.08	0.01
	Continuous		**0.77**	0.34	0.00	0.00
Abortion	Binned	10.0	0.90	**0.91**	0.14	0.01
	Continuous		**0.87**	0.83	0.02	0.02
Outcome variable						
Hours	Binned	7.8	**0.95**	0.91	0.00	0.00
	Continuous		**0.90**	0.86	0.00	0.00
Urate	Binned	7.9	**0.94**	0.91	0.03	0.02
	Continuous		**0.92**	0.82	0.01	0.01
Resample, $N = 200$	Binned	31.5	**0.92**	0.92	0.00	0.00
	Continuous		**0.92**	0.90	0.00	0.00
Resample, $N = 400$	Binned	64.6	**0.94**	0.90	0.00	0.00
	Continuous		**0.95**	0.92	0.00	0.00
Penn World Tables ($N = 111,\ T = 48$)						
Democracy	Binned	22.0	**0.93**	0.91	0.41	0.07
	Continuous		**0.85**	0.60	0.06	0.06
Education	Binned	13.9	**0.92**	0.84	0.56	0.39
	Continuous		**0.89**	0.42	0.33	0.33

Table A4—: Coverage results for nominal 95% coverage rates, with and without a slope term in the DGP. 500 replications, doses are grouped into $K = 5$ bins, $\bar{N}_1$ is the mean number of treated units across replications, and $T_{post} = 10$. cDiD is DiD with a continuous treatment from Callaway, Goodman-Bacon and Sant'Anna (2024*a*).

			Slope				No Slope			
Simulator	Dose Grouping	$\bar{N}_1$	SDID	SC	DiD	cDiD	SDID	SC	DiD	cDiD
Current Population Survey ($N = 50$,$T = 40$)										
Baseline	Binned	8.2	**0.93**	0.87	0.04	0.02	**0.93**	0.91	0.85	0.82
	Continuous		**0.91**	0.82	0.01	0.01	**0.93**	0.86	0.82	0.82
Assignment process										
Gun law	Binned	25.0	**0.54**	0.29	0.00	0.00	0.68	**0.85**	0.73	0.76
	Continuous		**0.36**	0.00	0.00	0.00	0.54	0.66	**0.79**	**0.79**
Abortion	Binned	17.9	0.86	**0.91**	0.05	0.00	0.96	0.76	**0.99**	**0.99**
	Continuous		**0.80**	0.64	0.00	0.00	0.96	0.20	**0.99**	**0.99**
Outcome variable										
Hours	Binned	7.9	**0.96**	0.92	0.00	0.00	0.95	0.98	0.98	**1.00**
	Continuous		**0.90**	0.87	0.00	0.00	0.94	0.98	**1.00**	**1.00**
Urate	Binned	8.2	**0.94**	0.89	0.03	0.02	**0.96**	0.95	0.90	0.85
	Continuous		**0.92**	0.80	0.01	0.01	**0.92**	**0.92**	0.83	0.83
Resample, $N = 200$	Binned	31.6	**0.92**	0.92	0.00	0.00	**0.92**	0.89	0.70	0.35
	Continuous		**0.92**	0.90	0.00	0.00	**0.90**	0.77	0.85	0.37
Resample, $N = 400$	Binned	64.6	**0.94**	0.90	0.00	0.00	**0.91**	0.90	0.65	0.09
	Continuous		**0.96**	0.92	0.00	0.00	**0.95**	0.78	0.90	0.08
Penn World Tables ($N = 111, T = 48$)										
Democracy	Binned	28.6	0.88	**0.92**	0.26	0.04	0.88	**0.92**	0.26	0.04
	Continuous		**0.83**	0.56	0.02	0.02	**0.83**	0.56	0.02	0.02
Education	Binned	13.9	**0.92**	0.84	0.56	0.39	**0.92**	0.84	0.56	0.39
	Continuous		**0.89**	0.42	0.33	0.33	**0.89**	0.42	0.33	0.33

Algorithm A1 Inference for $\hat{\tau}_c(d_c)$

Data: $\mathbf{Y}$, D_i, T_{pre}, estimator $\in \{\text{TWFE}, \text{SC}, \text{SDID}\}$, B.
Result: $\widehat{V}_c^{\text{block}}$, $\widehat{V}_c^{\text{jk}}$, $\widehat{V}_c^{\text{pl}}$ for all $c \in \mathbf{C}$.

Block Bootstrap

for $c \in \mathbf{C}$ **do**
  for $b = 1$ **to** B **do**
    Sample $|\mathcal{G}_c|$ units with replacement from $(\mathcal{G}_0 \cup \mathcal{G}_c, D)$. Discard draws with no treated units.
    Compute $\hat{\tau}_c^{(b)}$ using Algorithm 1 on the resampled sub-panel.
  end for
  $\widehat{V}_c^{\text{block}} = \frac{1}{B}\sum_{b=1}^{B}\big(\hat{\tau}_c^{(b)} - \frac{1}{B}\sum_{b=1}^{B}\hat{\tau}_c^{(b)}\big)^2$.
end for

Jackknife

for $c \in \mathbf{C}$ **do**
  On the full sub-panel $\mathcal{G}_0 \cup \mathcal{G}_c$, estimate baseline weights:

$$\hat{\Omega}_c = \begin{cases} \{\hat{\omega}_c,\, \hat{\lambda}\} & \text{if SDID} \\ \{\hat{\omega}_c\} & \text{if SC} \\ \{N_0^{-1}\mathbf{1},\, T_{\text{pre}}^{-1}\mathbf{1}\} & \text{if TWFE} \end{cases}$$

  and compute baseline estimate $\hat{\tau}_c$.
  Set $N_c = |\mathcal{G}_0 \cup \mathcal{G}_c|$.
  for $b = 1$ **to** N_c **do**
    Remove unit b from the sub-panel $(Y^{(-b)}, D^{(-b)})$.
    Re-specify weights $\hat{\Omega}_c^{(-b)}$ on the leave-one-out sub-panel.
    Compute $\hat{\tau}_c^{(-b)}$ using Algorithm 1 with set weights $\hat{\Omega}_c^{(-b)}$.
  end for
  Set $\widehat{V}_c^{\text{jk}} = \frac{N_c-1}{N_c}\sum_{b=1}^{N_c}\big(\hat{\tau}_c^{(-b)} - \hat{\tau}_c\big)^2$.
end for

Placebo

for $c \in \mathbf{C}$ **do**
  for $b = 1$ **to** B **do**
    Draw $|\mathcal{G}_c|$ units without replacement from $\mathcal{G}_0$, remove them from the donor pool and assign them treatment dose $\{D_i : i \notin \mathcal{G}_0\}$.
    Compute $\hat{\tau}_c^{(b)}$ using Algorithm 1
  end for
  Set $\widehat{V}_c^{\text{pl}} = \frac{1}{B}\sum_{b=1}^{B}\big(\hat{\tau}_c^{(b)} - \frac{1}{B}\sum_{b=1}^{B}\hat{\tau}_c^{(b)}\big)^2$.
end for

Algorithm A2 Block bootstrap inference for $\widehat{ACRT}^{\circ}$

Data: $\mathbf{Y}$, D_i, T_{pre}, B.
Result: Variance estimator $\widehat{V}^{\text{block}}_{ACRT^{\circ}}$.

1: **for** $b \leftarrow 1$ **to** B **do**
2: Construct a bootstrap dataset $(\mathbf{Y}^{(b)}, D_i^{(b)})$ by sampling N rows of $(\mathbf{Y}, D_i)$ with replacement.
3: **if** *any dose cohort* $\mathcal{G}_c$ *is empty in bootstrap sample* **then**
4: Discard resample and go to 2.
5: **end**
6: Compute $\widehat{ACRT}^{\circ}_{(b)}$ following Algorithm 1 and (A1) using $(\mathbf{Y}^{(b)}, D_i^{(b)})$.
7: **end for**
8: Define $\widehat{V}^{\text{block}}_{ACRT^{\circ}} = \dfrac{1}{B}\sum_{b=1}^{B}\Big(\widehat{ACRT}^{\circ}_{(b)} - \frac{1}{B}\sum_{b=1}^{B}\widehat{ACRT}^{\circ}_{(b)}\Big)^2$.

Algorithm A3 Jackknife inference for $\widehat{ACRT}^{\circ}$

Inputs: $\mathbf{Y}$, D_i, T_{pre}, $\widehat{ACRT}^{\circ}$.
Result: Variance estimator $\widehat{V}^{\text{jk}}_{ACRT^{\circ}}$.

1: **for** $i \leftarrow 1$ **to** N **do**
2: Compute $\widehat{ACRT}^{\circ}_{(-i)}$ following Algorithm 1 and (A1) using $(\mathbf{Y}^{(-i)}, D_i^{(-i)})$.
3: **end for**
4: Compute $\widehat{V}^{\text{jk}}_{ACRT^{\circ}} = (N-1)N^{-1}\sum_{i=1}^{N}\Big(\widehat{ACRT}^{\circ}_{(-i)} - \widehat{ACRT}^{\circ}\Big)^2$.

Algorithm A4 Placebo inference for $\widehat{ACRT}^{\circ}$

Inputs: $\mathbf{Y}_{co}$, D_i, B.
Result: Variance estimator $\widehat{V}^{\text{pl}}_{ACRT^{\circ}}$.

1: **for** $b \leftarrow 1$ **to** B **do**
2: Draw $N_{tr} := \sum_{c=1}^{C}|\mathcal{G}_c|$ units without replacement from $\mathcal{G}_0$, remove them from the donor pool and assign them dose $\{D_i : i \notin \mathcal{G}_0\}$.
3: Construct placebo treatment matrix $\mathbf{W}^{(b)}_{co}$
4: Compute $\widehat{ACRT}^{\circ}_{(b)}$ following Algorithm 1 and (A1) based on $(\mathbf{Y}_{co}, \mathbf{W}^{(b)}_{co})$
5: **end for**
6: Define $\widehat{V}^{\text{pl}}_{ACRT^{\circ}} = \dfrac{1}{B}\sum_{b=1}^{B}\Big(\widehat{ACRT}^{\circ}_{(b)} - \frac{1}{B}\sum_{b=1}^{B}\widehat{ACRT}^{\circ}_{(b)}\Big)^2$.

Second-Stage Design Conditions and Consistency

This appendix states sufficient conditions for second-stage estimation of the dose-response. The argument proceeds in three steps. First, I define the diagonal population dose-response curve and relate the cohort-specific first-stage targets to evaluations of this curve at the realized cohort doses. Second, I impose support conditions ensuring that the realized cohort doses become dense in the treated support and that their empirical distribution converges to a nondegenerate limit. Third, I strengthen the first-stage SDID results to uniform control across realized cohorts. This justifies second-stage estimation of both linear and nonparametric dose-response specifications, and implies consistency of the associated average slope summaries.

Let the treated dose support satisfy

$$\mathcal{D} = \{d_L\} \cup \mathcal{D}^+, \qquad \mathcal{D}^+ = [\underline{d}, \bar{d}] \subset \mathbb{R}_+, \qquad d_L < \underline{d} < \bar{d}.$$

Define the diagonal dose-response curve by

$$f(d) := \frac{1}{T_{\text{post}}} \sum_{t=T_{\text{pre}}+1}^{T} \mathbb{E}[Y_t(d) - Y_t(d_L) \mid D = d], \qquad d \in \mathcal{D}^+.$$

I assume that f is well-defined and continuously differentiable on $\mathcal{D}^+$. For each realized cohort dose $d_c \in \mathcal{D}^+$, the first-stage target satisfies

$$\tau_c = ATT(d_c \mid d_c) = f(d_c).$$

Assumption 7 (Cohort support)**.** *Let $d_1 < \cdots < d_C$ denote the realized treated cohort doses, and let*

$$\hat{H}_C := \frac{1}{C} \sum_{c=1}^{C} \delta_{d_c}$$

denote their empirical distribution. Assume:

(i) $C \to \infty$, $d_1 \to \underline{d}$, $d_C \to \bar{d}$, and $\max_{1 \le c < C}(d_{c+1} - d_c) \to 0$;

(ii) $\hat{H}_C \Rightarrow H_D$ for some probability distribution H_D with support $\mathcal{D}^+$.

Proposition 3 (Consequences of cohort support)**.** *Suppose Assumption 7 holds. Then, for every bounded continuous function $g : \mathcal{D}^+ \to \mathbb{R}$,*

$$\frac{1}{C} \sum_{c=1}^{C} g(d_c) \to \int_{\mathcal{D}^+} g(d)\, dH_D(d).$$

In particular,

$$\frac{1}{C}\sum_{c=1}^{C} f(d_c) \to \int_{\mathcal{D}^+} f(d)\, dH_D(d), \qquad \frac{1}{C}\sum_{c=1}^{C} f'(d_c) \to \int_{\mathcal{D}^+} f'(d)\, dH_D(d).$$

Proof. By Assumption 7(ii), $\hat{H}_C \Rightarrow H_D$. Hence, for every bounded continuous g,

$$\frac{1}{C}\sum_{c=1}^{C} g(d_c) = \int_{\mathcal{D}^+} g(d)\, d\hat{H}_C(d) \to \int_{\mathcal{D}^+} g(d)\, dH_D(d)$$

by the Portmanteau theorem. The stated conclusions follow by taking $g = f$ and $g = f'$, both of which are continuous on the compact set $\mathcal{D}^+$. □

Assumption 8 (Uniform first-stage concentration)**.** *There exist constants $K_1, K_2 > 0$ such that, for all sufficiently large samples, all $x \geq 0$, and all $c \in \mathbf{C}$,*

$$\Pr\left(\sqrt{|\mathcal{G}_c|\, T_{\text{post}}}\, |\hat{\tau}_c - \tau_c| > x\right) \leq K_1 e^{-K_2 x^2}.$$

Moreover,

$$\frac{\log C}{\big(\min_{1\leq c\leq C} |\mathcal{G}_c|\big) T_{\text{post}}} \to 0.$$

Theorem 3 (Uniform first-stage consistency)**.** *Suppose Assumption 8 holds. Then*

$$\max_{1\leq c\leq C} |\hat{\tau}_c - \tau_c| = o_p(1).$$

If, in addition, $\tau_c = f(d_c)$ for all realized cohort doses $d_c \in \mathcal{D}^+$, then

$$\max_{1\leq c\leq C} |\hat{\tau}_c - f(d_c)| = o_p(1).$$

Proof. Let

$$m_C := \big(\min_{1\leq c\leq C} |\mathcal{G}_c|\big) T_{\text{post}}.$$

Fix $\varepsilon > 0$. By the union bound and Assumption 8,

$$\Pr\left(\max_{1\leq c\leq C} |\hat{\tau}_c - \tau_c| > \varepsilon\right) \leq C K_1 \exp(-K_2 \varepsilon^2 m_C).$$

Since $\log C / m_C \to 0$, the right-hand side converges to zero. Therefore

$$\max_{1\leq c\leq C} |\hat{\tau}_c - \tau_c| = o_p(1).$$

The second claim follows immediately from $\tau_c = f(d_c)$. □

B1. Second-Stage Estimation

The first-stage SDID procedure delivers the cohort-level sample

$$\{(d_c, \hat{\tau}_c)\}_{c=1}^{C},$$

where $\hat{\tau}_c$ estimates the diagonal target

$$\tau_c = f(d_c) = ATT(d_c \mid d_c).$$

Under Assumptions 7–8, the second stage treats these cohort-level estimates as uniformly consistent evaluations of the underlying dose-response curve. Define the design-based average slope summary

$$S^{\circ} := \int_{\mathcal{D}^+} f'(d)\, dH_D(d).$$

If, in addition, $H_D = F_D(\cdot \mid D > d_L)$, then

$$S^{\circ} = ACRT^{\circ} = \mathbb{E}\big[f'(D) \mid D > d_L\big].$$

Linear specification

Assumption 9 (Exact linear diagonal dose-response)**.** *There exist $\alpha, \beta \in \mathbb{R}$ such that*

$$f(d) = \alpha + \beta d \qquad \textit{for all } d \in \mathcal{D}^+.$$

Define the linear estimator by

$$(\hat{\alpha}, \hat{\beta}) = \underset{(a,b)\in\mathbb{R}^2}{\arg\min} \sum_{c=1}^{C} (\hat{\tau}_c - a - b d_c)^2.$$

Theorem 4 (Consistency of the linear second stage)**.** *Suppose Assumptions 7, 8, and 9 hold. Then*

$$\hat{\alpha} \xrightarrow{p} \alpha \qquad \textit{and} \qquad \hat{\beta} \xrightarrow{p} \beta = S^{\circ}.$$

If, in addition, $H_D = F_D(\cdot \mid D > d_L)$, then

$$\hat{\beta} \xrightarrow{p} ACRT^{\circ}.$$

Proof. Write

$$\hat{\tau}_c = \alpha + \beta d_c + r_c, \qquad \max_{1\le c\le C} |r_c| = o_p(1),$$

where the uniform remainder statement follows from Theorem 3. Let

$$\bar{d} := \frac{1}{C}\sum_{c=1}^{C} d_c, \qquad \bar{\hat{\tau}} := \frac{1}{C}\sum_{c=1}^{C} \hat{\tau}_c.$$

Then the OLS slope admits the representation

$$\hat{\beta} = \frac{\sum_{c=1}^{C}(d_c - \bar{d})(\hat{\tau}_c - \bar{\hat{\tau}})}{\sum_{c=1}^{C}(d_c - \bar{d})^2} = \beta + \frac{\sum_{c=1}^{C}(d_c - \bar{d})(r_c - \bar{r})}{\sum_{c=1}^{C}(d_c - \bar{d})^2},$$

where $\bar{r} = C^{-1}\sum_{c=1}^{C} r_c$. Since the support of H_D is $\mathcal{D}^+$, the denominator is asymptotically proportional to C, while the numerator is bounded in absolute value by a constant multiple of $C \max_c |r_c|$. Hence

$$\hat{\beta} - \beta = o_p(1).$$

The intercept follows from

$$\hat{\alpha} = \bar{\hat{\tau}} - \hat{\beta}\,\bar{d}$$

together with $\bar{\hat{\tau}} \to \alpha + \beta \int d\, dH_D(d)$ and $\hat{\beta} \to \beta$. Under Assumption 9, $f'(d) = \beta$ on $\mathcal{D}^+$, so $S^\circ = \beta$. If $H_D = F_D(\cdot \mid D > d_L)$, then $S^\circ = ACRT^\circ$. □

The linear consistency result does not by itself justify conventional OLS standard errors, because the cohort-specific first-stage estimates share a common donor pool and therefore need not be cross-cohort independent. The next corollary records a compact sufficient condition for asymptotically valid heteroskedasticity-robust inference.

Assumption 10 (Additional conditions for analytic HC inference)**.** *The centered design is nondegenerate:*

$$\frac{1}{C}\sum_{c=1}^{C}(d_c - \bar{d})^2 \to Q_d \qquad \text{for some } Q_d \in (0, \infty).$$

No single cohort dominates the asymptotic linear score. Moreover, after aggregation through the centered design weights $(d_c - \bar{d})$, the shared-control component, oracle confounding bias, and deviation-from-oracle terms are asymptotically negligible, and HC0 residual replacement is valid.

Corollary 1 (Analytic HC inference for the linear slope)**.** *Suppose Assumptions 7, 8, 9, and 10 hold. Then*

$$\frac{\hat{\beta} - \beta}{\hat{V}_{\beta,\mathrm{HC0}}^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where $\hat{V}_{\beta,\mathrm{HC0}}$ denotes the heteroskedasticity-robust variance estimator from the second-stage regression. Hence the associated HC0 Wald interval is asymptotically valid for β, and therefore for S°, and for $ACRT^\circ$ whenever $H_D = F_D(\cdot \mid D > d_L)$.

Define the penalized estimator

$$\hat{f} = \arg\min_{g \in \mathcal{H}^m} \left\{ \frac{1}{C} \sum_{c=1}^{C} \big(\hat{\tau}_c - g(d_c)\big)^2 + \mu_C J(g) \right\}, \tag{A1}$$

where $\mathcal{H}^m$ is a class of sufficiently smooth functions, $J(g)$ is a smoothness penalty, and μ_C is a tuning parameter.

Assumption 11 (Nonparametric recovery)**.** *When applied to noiseless observations $\{(d_c, f(d_c))\}_{c=1}^{C}$, the estimator in* (A1) *consistently recovers both the curve and its derivative:*

$$\sup_{d \in \mathcal{D}^+} |\tilde{f}(d) - f(d)| \to 0, \qquad \sup_{d \in \mathcal{D}^+} |\tilde{f}'(d) - f'(d)| \to 0,$$

where $\tilde{f}$ denotes the infeasible estimator constructed from $\{(d_c, f(d_c))\}_{c=1}^{C}$. In addition, replacing $f(d_c)$ by $\hat{\tau}_c$ perturbs the fitted curve and its derivative by at most a vanishing amount uniformly over $\mathcal{D}^+$.

Define the corresponding average slope estimator by

$$\hat{S}^\circ := \frac{1}{C} \sum_{c=1}^{C} \hat{f}'(d_c).$$

Theorem 5 (Consistency of the nonparametric second stage)**.** *Suppose Assumptions 7, 8, and 11 hold. Then*

$$\sup_{d \in \mathcal{D}^+} |\hat{f}(d) - f(d)| = o_p(1), \qquad \sup_{d \in \mathcal{D}^+} |\hat{f}'(d) - f'(d)| = o_p(1),$$

and

$$\hat{S}^\circ \xrightarrow{p} S^\circ = \int_{\mathcal{D}^+} f'(d)\, dH_D(d).$$

If, in addition, $H_D = F_D(\cdot \mid D > d_L)$, then

$$\hat{S}^\circ \xrightarrow{p} ACRT^\circ.$$

Proof. By Assumption 11,

$$\sup_{d \in \mathcal{D}^+} |\hat{f}(d) - \tilde{f}(d)| = o_p(1), \qquad \sup_{d \in \mathcal{D}^+} |\hat{f}'(d) - \tilde{f}'(d)| = o_p(1),$$

and

$$\sup_{d \in \mathcal{D}^+} |\tilde{f}(d) - f(d)| \to 0, \qquad \sup_{d \in \mathcal{D}^+} |\tilde{f}'(d) - f'(d)| \to 0.$$

Hence

$$\sup_{d\in\mathcal{D}^+} |\hat{f}(d) - f(d)| = o_p(1), \qquad \sup_{d\in\mathcal{D}^+} |\hat{f}'(d) - f'(d)| = o_p(1).$$

Now write

$$\hat{S}^\circ - S^\circ = \frac{1}{C}\sum_{c=1}^{C}\left[\hat{f}'(d_c) - f'(d_c)\right] + \left[\frac{1}{C}\sum_{c=1}^{C} f'(d_c) - \int_{\mathcal{D}^+} f'(d)\, dH_D(d)\right].$$

The first term is bounded in absolute value by $\sup_d |\hat{f}'(d) - f'(d)| = o_p(1)$, and the second converges to zero by Proposition 3. Therefore

$$\hat{S}^\circ \xrightarrow{p} S^\circ.$$

If $H_D = F_D(\cdot \mid D > d_L)$, then $S^\circ = ACRT^\circ$. □