Semi-parametric Non-linear Difference-in-Differences for Policy Evaluation: New Evidence on the Entrepreneurial Impacts of State Business Tax Credits *

Spencer Sween†

January 4, 2026

Abstract

I develop semiparametric difference-in-differences estimators for discrete, bounded, and weakly positive outcomes under identification assumptions that impose parallel trends on outcome-consistent nonlinear link scales. The framework is motivated by applications in entrepreneurship, where outcomes such as startup formation, patenting, entry, and firm quality are sparse, bounded, or frequently zero, rendering common log transformations and linear parallel trends assumptions inappropriate and potentially misleading. The proposed estimators accommodate continuous treatment intensity and allow for flexible, high-dimensional covariate adjustment using modern machine learning methods. I apply the framework to re-examine the effects of state research and development tax credits using data from the Startup Cartography Project and the Kauffman Survey of Entrepreneurial Activity, building on Fazio, Guzman, and Stern (2020). In contrast to prior positive estimates obtained from linear difference-in-differences specifications with log-transformed outcomes, the outcome-consistent nonlinear estimates imply significant negative effects of R&D credits on new startup formation.

*Preliminary work. Please do not circulate.

†University of California, Santa Barbara. spencersween@ucsb.edu

1 Introduction

Difference-in-differences (DiD) designs are central to empirical work in economics because they deliver transparent identifying assumptions, interpretable estimands, and credible counterfactuals in policy evaluation. Figure X documents a sharp increase in the use of DiD methods in leading journals since 2010. Identification in these designs rests on a parallel trends assumption for untreated potential outcomes, which is most often imposed in levels and applied across a wide range of outcome types, including binary indicators, counts, and bounded indices. In practice, researchers either transform outcomes to obtain proportional interpretations or to accommodate skewness, or they impose linear parallel trends directly on outcomes whose support is discrete or limited. Both practices embed implicit scale restrictions that are consequential for identification and interpretation.

Recent work emphasizes that the content of the parallel trends assumption is inherently scale dependent. Parallel trends in levels generally do not imply parallel trends after nonlinear transformations, nor do transformed-outcome trends correspond to well-defined causal objects on the original outcome scale (CITE). When outcomes admit zeros, commonly used log-plus-one specifications may fail to define coherent counterfactuals (CITE). More broadly, linear trends imposed on discrete or bounded outcomes can generate counterfactual paths outside the support of the data. These concerns reflect a fundamental mismatch between the scale on which identification is imposed and the process governing outcome evolution. As a result, standard DiD implementations can produce estimates that are difficult to interpret and potentially misleading, even when pre-treatment trends appear stable.

A natural alternative is to impose parallel trends on a link scale that is consistent with the outcome's support and dynamics, such as the log-mean scale for count outcomes or the log-odds scale for binary outcomes. Recent methodological contributions formalize nonlinear parallel trends restrictions and develop estimators for binary and count outcomes under staggered adoption with binary treatment (CITE). While these approaches clarify identification in nonlinear settings, they typically rely on low-dimensional, parametric covariate structures and do not accommodate continuous treatment intensities.

This paper develops semi-parametric DiD estimators for discrete and weakly positive outcomes under nonlinear parallel trends assumptions. The proposed framework targets causal effects defined on outcome-consistent link scales, thereby preserving support restrictions and policy-relevant interpretations. The estimators allow for flexible adjustment for high-dimensional covariates using modern machine learning methods and accommodate continuous treatments. Identification is maintained under conditional nonlinear parallel trends, extending existing DiD results beyond binary treatments and parametric specifications. Standard linear DiD arises as a special case of the framework, but the methods are designed for settings in which proportional effects or probability changes are the primary objects of interest.

The methodological contributions are motivated by applications in entrepreneurship, a domain in which outcomes are naturally discrete or bounded and policies often vary in intensity. Measures of entrepreneurial activity include counts of new firm formation, patenting events, and financing outcomes, as well as binary indicators for entry, survival, and growth milestones. Public interventions, such as research and development tax credits, are implemented both discretely through enactment and continuously through variation in generosity. These features make entrepreneurship a natural setting for outcome-consistent nonlinear DiD, yet much of the existing empirical literature relies on ad hoc outcome transformations or linear parallel trends imposed on scale-inconsistent outcomes.

I apply the proposed methods to data from the Startup Cartography Project (CITE), which provides geographically and temporally detailed measures of entrepreneurial activity in the United States. I re-examine the effects of state research and development tax credits on startup formation and firm quality, following Fazio, Guzman, and Stern (2020). Using county- and ZIP-code-level data augmented with rich covariates capturing local economic conditions, I estimate effects under conditional nonlinear parallel trends on outcome-consistent link scales and allow treatment intensity to vary with tax credit generosity. I further extend the analysis to individual-level entry decisions using microdata from the Kauffman Survey of Entrepreneurial Activity, which permits an examination of extensive-margin responses at the household level and provides an independent validation of the aggregate results.

The results indicate large negative effects of research and development tax credits on startup formation, firm quality, and individual entry, in contrast to prior positive estimates obtained under linear two-way fixed effects specifications with log-plus-one outcomes.

Taken together, the paper develops a researcher-ready framework for DiD designs with discrete or bounded outcomes, non-binary treatments, and flexible covariate adjustment. While the empirical application focuses on entrepreneurship, the methods apply broadly to policy evaluations in which proportional effects and outcome support constraints are central, including trade flows under tariff changes, fertility responses to policy reforms, and public health interventions measured by incidence counts.

Roadmap. Section 2 reviews the related literature and introduces the data and replicated empirical setting. Section 3 defines the target parameters, and Section 4 presents the main identification results for commonly used empirical designs and treatment structures. Section 5 develops the semi-parametric estimators and discusses the trade-offs between linear and nonlinear specifications, along with placebo and specification checks. Section 6 presents extensions of the framework, including results for repeated cross-section designs. Section 7 applies the proposed methods to study the effects of research and development tax credits, revisiting Fazio, Guzman, and Stern (2020). Section 8 concludes.

2 Relevant Literature and Empirical Context

A growing methodological literature extends difference-in-differences designs beyond linear, additive specifications. This work formalizes nonlinear parallel trends restrictions that are consistent with the support of the outcome, such as parallel growth on the log-mean scale for count outcomes and parallel movements on the log-odds scale for binary outcomes, and develops estimators that preserve proportional or probability-based interpretations (CITE). Recent contributions clarify identification and estimands in settings with staggered treatment adoption and heterogeneous effects, consider multi-period designs with non-binary or repeated exposure, and establish valid inference under flexible covariate adjustment using orthogonal scores, sample splitting, and modern machine learning methods (CITE). Related work studies robustness to violations of standard DiD assumptions and proposes diagnostic and sensitivity tools for assessing identifying restrictions (CITE).

2.1 The Startup Cartography Project

The Startup Cartography Project (SCP) is a data infrastructure designed to measure entrepreneurial activity at high geographic and temporal resolution (CITE). The SCP links public business registration records to longitudinal firm identifiers and augments these records with early growth signals drawn from private and administrative sources. The resulting dataset provides near-population coverage of new business registrations, harmonizes entities across filing systems, and attaches information that proxies for growth orientation, including intellectual property activity, early hiring and payroll signals, external financing indicators, incorporation form, Delaware registration, and measures of online presence (CITE).

Formation measures count new business registrations within a place–time cell. These outcomes are weakly positive, highly skewed, and often sparse at fine geographic levels. Second, growth-event indicators record whether at least one milestone occurs within a given cell and period, such as an initial hire, a patent filing, or first external financing (CITE). Third, quality measures aggregate founding-time signals into indices that summarize the predicted probability of achieving a meaningful growth outcome within a fixed horizon, such as an initial public offering or acquisition.

The SCP can be merged with a rich set of regional covariates, including demographic characteristics, labor market conditions, industrial composition, higher education presence, infrastructure, and concurrent policy environments, as well as with denominators such as population or employment to construct rates or offsets (CITE). Many state-level entrepreneurship policies are introduced at different times and vary continuously in generosity. As a result, the SCP naturally supports designs that exploit both staggered timing and continuous treatment intensity. These features make it a particularly suitable setting for DiD analyses of both extensive and intensive margins under outcome-consistent identification assumptions.

2.2 State R&D Tax Credits and Prior Evidence

Fazio, Guzman, and Stern (2020) study the relationship between state research and development tax credits and entrepreneurial activity using outcomes constructed from the SCP (CITE). The authors assemble a county-year panel that combines SCP measures of entry, early growth signals, and quality indices with state-level measures of R&D tax-credit adoption and generosity. Identification relies on difference-in-differences with two-way fixed effects, exploiting staggered enactment across states and cross-state variation in statutory parameters such as credit rates, bases, caps, and carryforward provisions (CITE).

The analysis considers formation counts, indicators for early growth milestones, and composite quality measures. To obtain proportional interpretations and to address skewness, the study applies log-plus-one transformations to count outcomes and analogous transformations to bounded measures (CITE). The main findings indicate that the introduction of state R&D tax credits is associated with increases in the quantity of entrepreneurial activity and, over longer horizons, in quality-adjusted measures, while other investment incentives are found to have weaker or negative associations with quality-adjusted outcomes (CITE).

3 Setup

We fix ideas by considering a two-period panel with n independent and identically distributed units indexed by $i = 1, \dots, n$. Each unit is drawn from a population distribution $\mathcal{P}$ and is observed in both periods. The observed data for unit i are

$$W_i = (Y_{i0}, Y_{i1}, D_{i1}, X_i) \overset{iid}{\sim} \mathcal{P},$$

where Y_{it} denotes the outcome in period $t \in \{0, 1\}$, $D_{i1} \in \mathbb{R}_{\geq 0}$ denotes post-period treatment exposure (which may be binary, multi-valued, or continuous), and X_i is a vector of pre-treatment covariates that may be high-dimensional. Throughout, we assume no unit is treated in the pre period, so $D_{i0} = 0$ for all i.

Let $Y_{it}(d)$ denote the potential outcome for unit i in period t under post-period treatment level d. We impose no anticipation:

$$Y_{i0}(d) = Y_{i0}(0) \quad \text{for all } d,$$

and assume SUTVA and consistency:

$$Y_{it} = Y_{it}(D_{i1}).$$

The untreated potential outcome $Y_{it}(0)$ represents the outcome that would be observed in period t absent treatment. Define the unit- and time-specific treatment effect as

$$\tau_{it}(d) = Y_{it}(d) - Y_{it}(0), \qquad \tau_{it}(0) = 0.$$

3.1 Target Parameters

We consider target parameters that remain well-defined for discrete, bounded, and weakly positive outcomes, and that accommodate both binary and continuously varying treatment exposure. Unless otherwise noted, Y_{it} denotes the outcome on its natural scale.

Binary treatment. When treatment is binary, $D_{i1} \in \{0, 1\}$, the canonical estimand is the level average treatment effect on the treated (ATT) in the post period,

$$ATT \equiv \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid D_{i1} = 1] \,. \tag{1}$$

This estimand is defined in outcome levels and applies regardless of whether the outcome is continuous, discrete, or bounded, provided expectations are well defined.

In many empirical applications, researchers seek proportional interpretations. These interpretations are typically motivated by modeling the logarithm of an underlying nonnegative outcome and imposing parallel trends in log outcomes. Let Y_{it} denote the outcome on its natural scale. The proportional effect implied by log-scale parallel trends is

$$ATT_{\%} \equiv \mathbb{E}\left[\frac{Y_{i1}(1) - Y_{i1}(0)}{Y_{i1}(0)} \;\middle|\; D_{i1} = 1\right], \tag{2}$$

which is undefined whenever $Y_{i1}(0) = 0$. In practice, applied researchers have historically conducted log-plus-one outcome transformations and argued for identification under parallel trends on the log-plus-one scale. However, Roth, 2022 that such an approach fails to return well-defined proportional causal effects.

In this paper, for proportional effects with weakly positive outcomes, we instead consider a weighted proportional estimand that remains well defined in the presence of zeros,

$$ATT_{\omega,\%} \equiv \mathbb{E}\left[\omega(W_i, Y_{i1}(0))\, \frac{Y_{i1}(1) - Y_{i1}(0)}{Y_{i1}(0)} \;\middle|\; D_{i1} = 1\right], \qquad \omega(W_i, 0) = 0. \tag{3}$$

The weighting function $\omega(\cdot)$ determines how unit-level proportional changes are aggregated and ensures that the estimand is a well-defined causal summary.

A leading example studied in recent work normalizes by the treated group's untreated mean outcome (see Roth, 2022; Wooldridge, 2023). In particular, one may choose

$$\omega(W_i, Y_{i1}(0)) = \frac{Y_{i1}(0)}{\mathbb{E}[Y_{i1}(0) \mid D_{i1} = 1]}.$$

Under this choice, $ATT_{\omega,\%}$ aggregates proportional changes relative to the treated group's average untreated outcome and admits a growth-rate interpretation. This estimand coincides with the parameter recovered by Poisson maximum likelihood estimators under nonlinear parallel trends for count outcomes without covariates (Wooldridge, 2023).

Continuous treatment. We also consider target parameters for designs with a continuous treatment. When treatment varies continuously, $D_{i1} \in \mathbb{R}_{\geq 0}$, a natural object of interest is the dose-specific average treatment effect,

$$ATT_D(d) \equiv \mathbb{E}[Y_{i1}(d) - Y_{i1}(0) \mid D_{i1} = d], \tag{4}$$

defined for d in the support of D_{i1}. One challenge with focusing on this parameter is the need to estimate density functions, and with covariates *conditional* density functions, which can be challenging. Alternatively, as a summary measure of the exposure–response relationship, we may consider an average slope effect among treated units, as in Chaisemartin et al. 2024:

$$ATT_S \equiv \mathbb{E}\left[\frac{Y_{i1}(D_{i1}) - Y_{i1}(0)}{D_{i1}} \,\middle|\, D_{i1} > 0\right]. \tag{5}$$

This parameter captures a linear approximation to the average marginal effect of a one-unit increase in treatment intensity, averaged over all exposure levels in the treated group.

When proportional interpretations are desired, motivated by log-scale modeling of non-negative outcomes, we analogously define a weighted proportional slope estimand,

$$ATT_{\omega,S,\%} \equiv \mathbb{E}\left[\frac{1}{D_{i1}}\,\omega(W_i, Y_{i1}(0))\,\frac{Y_{i1}(D_{i1}) - Y_{i1}(0)}{Y_{i1}(0)}\;\middle|\; D_{i1} > 0\right], \qquad \omega(W_i, 0) = 0. \tag{6}$$

4 Identification under Nonlinear Parallel Trends

This section establishes identification of the target parameters introduced in Section 3. Identification is based on a nonlinear extension of the parallel trends assumption, imposed on an outcome-consistent link scale. Let $\mathcal{L}(\cdot)$ denote a strictly monotone link function applied to conditional mean outcomes, with inverse $\mathcal{L}^{-1}(\cdot)$. Examples include the logarithmic link for count outcomes and the logit link for binary outcomes. The standard linear difference-in-differences assumption corresponds to the identity link, $\mathcal{L}(m) = m$. Throughout, we maintain consistency and no anticipation, so that $Y_{it} = Y_{it}(D_{i1})$ and $Y_{i0}(d) = Y_{i0}(0)$ for all d.

Non-Linear Parallel Trends, Binary Treatment. When treatment is binary, $D_{i1} \in \{0, 1\}$, identification relies on the following nonlinear parallel trends condition:

Assumption 1. For some link function $\mathcal{L}$,

$$\begin{aligned} &\mathcal{L}\Big(\mathbb{E}[Y_{i1}(0) \mid D_{i1} = 1, X_i]\Big) - \mathcal{L}\Big(\mathbb{E}[Y_{i0}(0) \mid D_{i1} = 1, X_i]\Big) \\ &= \mathcal{L}\Big(\mathbb{E}[Y_{i1}(0) \mid D_{i1} = 0, X_i]\Big) - \mathcal{L}\Big(\mathbb{E}[Y_{i0}(0) \mid D_{i1} = 0, X_i]\Big). \end{aligned} \tag{7}$$

This assumption requires that, absent treatment, the evolution of expected outcomes is parallel on the link scale for treated and untreated units, conditional on covariates.

Under (7), the treated group's counterfactual mean outcome is identified as

$$\Delta_{i1}(X_i) = \mathcal{L}^{-1}\Big(\mathcal{L}\Big(\mathbb{E}[Y_{i0} \mid D_{i1} = 1, X_i]\Big) + \mathcal{L}\Big(\mathbb{E}[Y_{i1} \mid D_{i1} = 0, X_i]\Big) - \mathcal{L}\Big(\mathbb{E}[Y_{i0} \mid D_{i1} = 0, X_i]\Big)\Big),$$

so that $\Delta_{i1}(X_i) = \mathbb{E}[Y_{i1}(0) \mid D_{i1} = 1, X_i]$.

Under Assumption 1, the level average treatment effect on the treated is given by

$$ATT = \mathbb{E}[Y_{i1} - \Delta_{i1}(X_i) \mid D_{i1} = 1].$$

To obtain proportional effects that remain well defined when outcomes are discrete or weakly positive, we fix the weighting function to

$$\omega(W_i, Y_{i1}(0)) \equiv \frac{Y_{i1}(0)}{\mathbb{E}[Y_{i1}(0) \mid D_{i1} = 1, X_i]},$$

which downweights units with small predicted average untreated outcomes. While other weighting functions may be possible to define here, the inverse conditional average denominator is an attractive choice, particularly if the covariate set is rich enough, so that, under the non-linear identification conditions, the denominator amounts to a high-quality prediction of the (unobserved) untreated potential outcome, implying closer-to-unit weights on the proportional effects. Under this choice, the weighted proportional ATT is expressed as

$$ATT_{\omega,\%} = \mathbb{E}\left[\frac{Y_{i1}}{\Delta_{i1}(X_i)} - 1 \;\middle|\; D_{i1} = 1\right].$$

Non-Linear Parallel Trends, Continuous Treatment. When treatment is continuous, $D_{i1} \in \mathbb{R}_{\geq 0}$, identification relies on the following extension.

Assumption 2. Given $\mathcal{L}$, and for all d in the support of D_{i1},

$$\begin{aligned}
&\mathcal{L}\Big(\mathbb{E}[Y_{i1}(0) \mid D_{i1} = d, X_i]\Big) - \mathcal{L}\Big(\mathbb{E}[Y_{i0}(0) \mid D_{i1} = d, X_i]\Big) \\
&= \mathcal{L}\Big(\mathbb{E}[Y_{i1}(0) \mid D_{i1} = 0, X_i]\Big) - \mathcal{L}\Big(\mathbb{E}[Y_{i0}(0) \mid D_{i1} = 0, X_i]\Big). \qquad (8)
\end{aligned}$$

This condition requires that, conditional on covariates, link-scale changes in untreated potential outcomes are invariant to treatment intensity.

Under (8), the counterfactual mean outcome for a unit exposed at level d is

$$\Delta_{i1}(X_i, d) = \mathcal{L}^{-1}\Big(\mathcal{L}\Big(\mathbb{E}[Y_{i0} \mid D_{i1} = d, X_i]\Big) + \mathcal{L}\Big(\mathbb{E}[Y_{i1} \mid D_{i1} = 0, X_i]\Big) - \mathcal{L}\Big(\mathbb{E}[Y_{i0} \mid D_{i1} = 0, X_i]\Big)\Big),$$

so that $\Delta_{i1}(X_i, d) = \mathbb{E}[Y_{i1}(0) \mid D_{i1} = d, X_i]$.

The dose-specific average treatment effect is therefore

$$ATT_D(d) = \mathbb{E}[Y_{i1} - \Delta_{i1}(X_i, d) \mid D_{i1} = d].$$

Meanwhile, the average slope estimand is

$$ATT_S = \mathbb{E}\left[\frac{Y_{i1} - \Delta_{i1}(X_i, D_{i1})}{D_{i1}} \;\middle|\; D_{i1} > 0\right].$$

Finally, fixing the weighting function to

$$\omega(W_i, Y_{i1}(0)) \equiv \frac{Y_{i1}(0)}{\mathbb{E}[Y_{i1}(0) \mid D_{i1}, X_i]},$$

the weighted proportional slope estimand is identified as

$$ATT_{\omega,S,\%} = \mathbb{E}\left[\frac{1}{D_{i1}}\left(\frac{Y_{i1}}{\Delta_{i1}(X_i, D_{i1})} - 1\right) \,\middle|\, D_{i1} > 0\right].$$

5 Semiparametric Estimation and Inference

This section develops semiparametric estimators and inference procedures for the parameters identified in Section 4. We begin with a two-period panel and binary treatment, where the estimands of interest are identified as nonlinear, globally defined functionals of several conditional mean functions. This formulation nests the level, proportional, and weighted proportional effects introduced earlier. We then outline how the same estimation logic extends to continuous treatments, multi-period panels, and repeated cross-sections.

Estimation throughout proceeds in two general steps. First, the relevant conditional mean functions are estimated flexibly using nonparametric or machine learning methods. Second, these estimates are combined using Neyman-orthogonal score functions and cross-fitting, yielding root-n consistent and asymptotically normal estimators under standard regularity conditions. The construction is based on the general framework of Chernozhukov et al., 2021.

5.1 Binary Treatments

Let $W_i = (Y_{i0}, Y_{i1}, D_i, X_i)$ denote observed data for unit i, where $D_i \in \{0, 1\}$ indicates treatment status in the post period and X_i denotes covariates. Let $\pi = \mathbb{P}(D_i = 1)$. For

$t \in \{0,1\}$ and $d \in \{0,1\}$, define the conditional mean functions

$$\mu_t^d(x) \equiv \mathbb{E}[Y_{it} \mid D_i = d, X_i = x].$$

Let $\mathcal{L}(\cdot)$ be a strictly monotone link function with inverse $\mathcal{L}^{-1}(\cdot)$. Assume Assumption 1 holds given $\mathcal{L}$ so that the treated group's counterfactual mean outcome is identified as

$$\Delta_1(X_i) = \mathbb{E}[Y_{i1}(0) \mid D_i = 1, X_i] = \mathcal{L}^{-1}\Big(\mathcal{L}(\mu_0^1(X_i)) + \mathcal{L}(\mu_1^0(X_i)) - \mathcal{L}(\mu_0^0(X_i))\Big).$$

Target Functionals. Let $H(W_i, \boldsymbol{\mu}; \mathcal{L})$ denote a smooth, globally defined functional of the data and nuisance functions. Given $\mathcal{L}$, a generic target parameter is

$$\theta_0 = \mathbb{E}[H(W_i, \boldsymbol{\mu}; \mathcal{L})]\,,$$

for some smooth transformation of the data and conditional mean functions, H.

This formulation nests all target parameters introduced in Section 3. For example, the weighted proportional average treatment effect on the treated under the link $\mathcal{L}$ is

$$\begin{aligned}
ATT_{\omega,\%} &= \mathbb{E}\left[\frac{Y_{i1}}{\Delta_1(X_i)} - 1 \,\middle|\, D_i = 1\right] \\
&= \mathbb{E}\left[\frac{D_i}{\pi}\left(\frac{\mu_1^1(X_i)}{\mathcal{L}^{-1}\big(\mathcal{L}(\mu_0^1(X_i)) + \mathcal{L}(\mu_1^0(X_i)) - \mathcal{L}(\mu_0^0(X_i))\big)} - 1\right)\right] \\
&= \mathbb{E}\left[H(X_i, D_i, \pi, \boldsymbol{\mu}; \mathcal{L})\right].
\end{aligned}$$

Notice the unconditional propensity score, π, can be treated as fixed in constructing an estimator, where we may then account for estimation of π in the final influence function by applying the Delta method.

Neyman Orthogonal Estimators. To obtain estimators that are robust to first-stage estimation error and facilitate valid inference after (potentially machine-learning-assisted) nuisance function estimation, we construct Neyman-orthogonal scores. Let $\psi(W_i; \theta, g)$ denote a score function satisfying

$$\mathbb{E}[\psi(W_i; \theta_0, g_0)] = 0, \qquad \frac{d}{d\eta}\mathbb{E}[\psi(W_i; \theta_0, g_0 + \eta\tau)]\big|_{\eta=0} = 0$$

where g_0 collects the true functions, and such that the Gateaux derivative vanishes at g_0.

We follow Chernozhukov et al., 2021, which provides a general construction of orthogonal influence functions for nonlinear functionals of multiple regression objects. Under regularity conditions, the influence function admits the representation

$$\psi(W_i; \theta_0, g_0) = H(W_i, \boldsymbol{\mu}; \mathcal{L}) - \theta_0 + \sum_{t=0}^{1}\sum_{d=0}^{1} \alpha_t^d(X_i)\mathbb{1}(D_i = d)\big(Y_{it} - \mu_t^d(X_i)\big),$$

where $\alpha_t^d(\cdot)$ are Riesz representers that debias first-stage estimation error.

Each Riesz Representer α_t^d is characterized as the minimizer of

$$\alpha_t^d = \arg\min_a \; \mathbb{E}\left[-2\,\frac{\partial}{\partial\eta} H\big(W_i, \boldsymbol{\mu} + \eta\,\mathbb{1}_{td}a(X_i); \mathcal{L}\big)\bigg|_{\eta=0} + \mathbb{1}(D_i = d)a(X_i)^2\right],$$

where $\mathbb{1}_{td}$ denotes a perturbation of the (t, d) component of $\boldsymbol{\mu}$. In practice, these terms are obtained via automatic differentiation and non-parametric regression.

Example – Closed-Form Score for Log-Mean Parallel Trends. For weakly positive count outcomes, a natural link function is $\mathcal{L}(x) = \log x$. In this case,

$$\Delta_1(x) = \exp\big(\log \mu_0^1(x) + \log \mu_1^0(x) - \log \mu_0^0(x)\big),$$

and the orthogonal score simplifies substantially. For illustration purposes, that we will employ in the empirical application, consider $\theta = ATT_{\omega,\%}$ and define $g(x) = (\mu_1^1(x), \mu_0^1(x), \mu_1^0(x), \mu_0^0(x))$. Then, the Neyman-orthogonal influence function (assuming π is known) is

$$\begin{aligned}\psi(W_i;\theta,g) = \frac{D_i}{\pi}\Bigg[&\frac{\mu_1^1(X_i)}{\Delta_1(X_i)} - 1 - \theta + \frac{Y_{i1} - \mu_1^1(X_i)}{\Delta_1(X_i)} - \frac{\mu_1^1(X_i)}{\Delta_1(X_i)}\frac{Y_{i0} - \mu_0^1(X_i)}{\mu_0^1(X_i)} \\ &- \frac{\mu_1^1(X_i)}{\Delta_1(X_i)}\frac{(1-D_i)(Y_{i1} - \mu_1^0(X_i))}{(1-\pi)\mu_1^0(X_i)} + \frac{\mu_1^1(X_i)}{\Delta_1(X_i)}\frac{(1-D_i)(Y_{i0} - \mu_0^0(X_i))}{(1-\pi)\mu_0^0(X_i)}\Bigg].\end{aligned} \tag{9}$$

General Implementation For a general parameter of interest with a binary treatment, estimation proceeds as follows:

1. Learn the conditional mean functions $\mu_t^d(x)$ using flexible regression or machine learning methods using cross-fitting with the squared loss function.
2. Compute the directional derivatives of the target functional with respect to each conditional mean, then minimize the Riesz Representer loss, again with cross-fitting.
3. Construct the orthogonal score on held-out data from the cross-fitting steps.

4. Obtain $\widehat{\theta}$ by solving the sample moment condition

$$\frac{1}{n}\sum_{i=1}^{n}\psi(W_i,\widehat{\theta},\widehat{g}) = 0.$$

5. Compute the asymptotic variance estimator as

$$\widehat{V}(\widehat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\psi(W_i,\widehat{\theta},\widehat{g})\psi(W_i,\widehat{\theta},\widehat{g})^{\mathrm{T}}$$

Under standard overlap and rate conditions, $\widehat{\theta}$ is root-n consistent and asymptotically normal. Standard errors are computed using the empirical variance of the cross-fitted influence function, and confidence intervals may be formed using either asymptotic normal approximations or a multiplier bootstrap.

5.2 Continuous Treatments

[To be added...]

5.3 Multiple Periods and Staggered Adoption

[To be added...]

5.4 Repeated Cross Sections

[To be added...]

6 Extensions.

[To be added...]

6.1 Benchmarking Intensive-Margin Effects for Count Data

[To be added...]

6.2 Generalized Structural Parallel Trend Assumptions

[To be added...]

Example – Hazard Modeling. *[To be added...]*

Example – Sample Selection. *[To be added...]*

Example – Discrete Choice Models. *[To be added...]*

Example – Compositional Difference-in-Differences. *[To be added...]*

7 Empirical Application

This section re-examines the effects of state research and development (R&D) tax credits on entrepreneurial activity, revisiting the analysis of Fazio et al., 2020. The application serves two purposes. First, it benchmarks the proposed nonlinear difference-in-differences framework against widely used estimators in the applied literature. Second, it illustrates how outcome-consistent identification and estimation can materially affect both magnitudes and interpretation in a prominent policy setting. I closely align the sample construction, geographic units, outcome definitions, and policy measures with the original study. Holding these features fixed, I contrast estimates obtained under log-transformed-outcome linear two-way fixed-effects (TWFE) designs with alternative approaches that explicitly account for staggered treatment timing and the natural support of the raw outcomes.

7.1 Data and Empirical Strategy

Table **??** reports summary statistics for the main outcomes, policy variables, and covariates. The analysis uses a county-level panel constructed from the Startup Cartography Project and the Panel Database of Incentives and Taxes. Outcomes capture three dimensions of entrepreneurial activity: counts of new firm formation, indicators for early growth-oriented milestones such as IPOs and acquisitions, and a predictive measures of startup quality. The treatment variable, state R&D tax credit exposure, is measured using the PDIT time series for each state that record both the timing of adoption and generosity of the credits. To control for local economic conditions, I augment the SCP data with county-level population and labor market measures from the U.S. Census Bureau and the Quarterly Census of Employment and

Wages (QCEW), including population counts, employment-to-population ratios, earnings, and industry employment shares by county measured in 1990.

The empirical strategy proceeds in three steps. First, I replicate the original two-way fixed-effects event studies and post-adoption estimates using the authors' outcome transformations—log-plus-one for counts and analogous transformations for bounded outcomes—along with the reported control variables and clustering scheme. Figure **??** plots the resulting TWFE event-study coefficients. For completeness, I also estimate the same TWFE specifications using outcomes in levels, highlighting the sensitivity of the results to scale choice. Second, to address biases arising from staggered treatment adoption, I re-estimate the analysis using the group–time average treatment effect estimator of Callaway and Sant'Anna, 2021. Figure **??** reports the Callaway–Sant'Anna event-study coefficients for transformed outcomes, while Figure **??** reports the corresponding estimates in levels. Third, I estimate outcome-consistent specifications based on nonlinear parallel trends. For the count outcomes (e.g., the number of startups), I impose the parallel trend assumption using a log-link function, so that conditional means are assumed to follow a parallel growth rate between the treated and never-treated groups in the untreated state.

7.2 Replication and Benchmarking Results

The TWFE replication closely reproduces the qualitative event-study patterns and post-adoption estimates reported in Fazio et al., 2020. Figure **??** shows positive post-treatment coefficients following adoption, consistent with the original paper's findings. Re-estimating the design using the Callaway and Sant'Anna estimator yields event-study profiles that are similar in timing but differ in magnitude. Figure **??** shows that, while the transformed-outcome estimates remain positive, they are attenuated relative to TWFE. Figure **??** demonstrates that when outcomes are analyzed in levels under linear parallel trends, the implied treatment effects are negative and statistically significant. Table **??** summarizes these contrasts numerically, comparing TWFE and Callaway–Sant'Anna estimates across logged and level outcomes.

7.3 Nonlinear Parallel Trends Results

I next turn to estimates obtained under nonlinear parallel trends assumptions. Figure **??** plots event-study coefficients based on Wooldridge, 2023, which imposes parallel trends on the log-mean scale for untransformed startup counts. The resulting estimates indicate large and statistically significant negative effects of R&D tax credits on startup formation.

Figures **??** and **??** present event-study estimates from the proposed nonlinear DiD framework, targeting the weighted proportional estimand $ATT_{\omega,\%}$. Figure **??** reports estimates without covariates, isolating the identifying content of the nonlinear parallel trends assumption. Figure **??** augments this specification with high-dimensional covariates derived from Census and QCEW data and estimated using regularized regression. The inclusion of covariates improves precision but does not materially alter the qualitative pattern of effects.

Across specifications, the nonlinear DiD estimates imply that state R&D tax credits reduce startup formation by approximately 10% per year. This contrasts sharply with the linear DiD estimates based on log-plus-one-transformed outcomes, which the replication exercise shows to be positive at approximately +9% (Table **??**). Table **??** reports these results. Taken together, these results underscore the importance of aligning identifying assumptions with the natural support and dynamics of the outcome. When parallel trends are imposed on an outcome-consistent scale, the estimated effects of state R&D tax credits on entrepreneurial entry are economically meaningful, robust across specifications, and opposite in sign to those obtained under commonly used transformed-outcome linear DiD designs.

8 Conclusion

This paper develops a semiparametric difference-in-differences framework for settings with discrete, bounded, or weakly positive outcomes and treatments that vary in intensity. By formulating identification on outcome-consistent link scales, the framework respects support restrictions and delivers causal estimands with clear policy interpretation. Drawing on recent advances in debiased machine learning, we propose estimators that accommodate flexible machine learning methods for nuisance estimation and support inference on a broad class of parameters, including those involving binary and continuous treatments.

Applied to state R&D tax credits, the framework reveals that commonly used transformed-outcome linear DiD designs can mask substantial sensitivity to scale and weighting. In contrast to prior positive findings, outcome-consistent nonlinear estimates indicate negative effects on startup formation and related entrepreneurial outcomes. More broadly, the results underscore the importance of aligning identifying assumptions with the natural scale of the data when the measures are weakly positive, discrete, or bounded.

Outside of empirical entrepreneurship applications, the methods developed here apply broadly to policy evaluations involving counts, rates, and probabilities, including trade flows, fertility responses, and public health interventions. A forthcoming companion `R` package, `nldid`, will facilitate implementation in broader applied work.

References

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

Chernozhukov, V., Newey, W. K., Quintas-Martinez, V., & Syrgkanis, V. (2021). Automatic debiased machine learning via riesz regression. *arXiv preprint*, *arXiv:2104.14737*. https://arxiv.org/abs/2104.14737

Fazio, C., Guzman, J., & Stern, S. (2020). The impact of state-level r&d tax credits on the quantity and quality of entrepreneurship. *Economic Development Quarterly*, *34*(2), 188–208.

Roth, J. (2022). Pretest with caution: Event-Study estimates after testing for Parallel Trends. *American Economic Review: Insights*, *4*(3), 305–322. https://doi.org/10.1257/aeri.20210472

Wooldridge, J. M. (2023). Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*, *26*(3), C31–C66.

9 Figures and Tables

10 Appendix

[To be added...]