Sensitivity Analysis for Difference-in-Discontinuities Designs

Yiwei Sun *

October, 2026

Abstract

Difference-in-discontinuities (DiDC) designs identify the treatment effect of a policy introduced at a cutoff that also determines exposure to a pre-existing confounding policy by comparing the post- and pre-period regression discontinuities. The canonical design assumes that the confounding discontinuity remains constant over time, which may be difficult to justify in practice. This paper develops an identification and inference framework when this assumption may be violated. Following the sensitivity framework developed in the difference-in-differences literature, we leverage observed discontinuities from multiple pre periods to inform restrictions on the evolution of the counterfactual discontinuities after the introduction of the policy. The resulting identified sets nest the conventional point-identified estimand as a special case. We use an existing joint asymptotic normality result for local-polynomial RD estimators and the consistency of the full cross-period covariance estimator to justify confidence sets based on moment inequality tests in this nonparametric DiDC setting. These sets provide pointwise asymptotic coverage under explicit regularity, bias, bandwidth, and sensitivity class conditions. Simulations show that, when confounding varies over time, the proposed procedure has better coverage probability relative to conventional DiDC inference.

Keywords: Difference-in-discontinuities, regression discontinuity design, difference-in-differences, sensitivity analysis, partial identification, inference

1 Introduction

Difference-in-differences (DiD) and regression discontinuity (RD) designs are among the most popular empirical strategies for causal inference using quasi-experimental data. They

*Department of Economics, Cornell University. Email: ys556@cornell.edu

exploit different kinds of variation in the data to identify the treatment effect of interest. DiD designs exploit variation across treatment groups over time by comparing the change in outcomes for the treated units with the change for the control units. Meanwhile, RD designs exploit cross-sectional variation in treatment assignment generated by a known threshold. At the same time, there are relevant empirical settings that may contain both forms of variation. The difference-in-discontinuities design (DiDC) aims to address limitations of both methods in such settings by combining both temporal and discontinuity-based sources of variation for program evaluation.

One of the canonical applications of DiDC designs was motivated by Grembi et al. (2016) to study the effect of relaxing fiscal rules on debt accumulation at the local government level in Italy, where whether a municipality is subject to the relaxation depends on a population threshold. The standard RD identification strategy fails in this setting because there is another confounding policy, the salary of the mayor, that changes at the same population cutoff. At the same time, DiD designs are questionable because large and small municipalities are typically subject to different outcome dynamics in public policy, which violates the parallel trend assumption. Eggers et al. (2018) make the same point by documenting how pervasive compound treatment and sorting are in population-threshold RD designs. These two papers both suggested using difference-in-discontinuities designs as the alternative empirical strategy. DiDC designs address this problem by subtracting a pre-treatment discontinuity from the post-treatment discontinuity. The difference identifies the treatment effect of interest if the confounding discontinuity remains constant over time, under the maintained RD identifying conditions. However, a constant confounding effect may be too strong in practice and hard to justify. The empirical problem is how we can still make credible causal conclusions about the treatment effect of interest when the validity of a constant confounding effect is questionable.

This paper develops a formal sensitivity analysis framework for identification and inference in sharp DiDC designs that leverages the structure of observing regression discontinuities over multiple time periods. We generalize the DiDC estimands by differencing the per-period discontinuity with respect to a baseline pre-treatment period and decompose them as the treatment effect of interest plus the deviation from a constant confounding effect. This representation connects DiDC to the sensitivity framework of Rambachan and Roth (2023) that studies relaxation of the parallel trend assumption in DiD and related methods. We can leverage the full pre-treatment path of the observed deviations to impose restrictions that inform the counterfactual evolution of the confounding effect after treatment. This framework yields identified sets of the treatment effects of interest in the post-periods while allowing researchers to exploit information in all available time periods and assess how sensitive the

conclusions are with respect to assumption violations. The framework nests the conventional constant-confounding DiDC estimand and linear extrapolation of the confounding path as special cases.

For estimation, we follow the standard practice in the RD literature by Calonico et al. (2014) and use local-polynomial regression with bias correction to obtain the vector of period-specific RD estimates and its full covariance matrix. Additionally, we show that the Rambachan and Roth (2023) conditional and hybrid inference procedure delivers confidence sets with asymptotic size control for treatment effects over the maintained class of confounding paths for each fixed data generating process in this nonparametric setting. To show this, we leverage the asymptotic joint normality of the vector of period-specific local-polynomial RD estimators and the consistency of the full cross-period covariance estimator result by Fusejima et al. (2025), under a fixed number of periods, a common time-invariant running variable, and explicit regularity, bias, and bandwidth conditions. The resulting confidence sets account for both statistical uncertainty in estimating the true pre-treatment discontinuity and identification uncertainty from not knowing exactly how to extrapolate the post-treatment discontinuity (Rambachan and Roth, 2023).

The distinction from existing theory for DiDC methods highlights how the pre-treatment discontinuities enter identification and inference under the sensitivity analysis framework. Existing approaches relax constant confounding through bounded-variation restrictions using ad-hoc constants (Picchetti et al., 2026) or use multiple periods to identify effects under assumptions of specific functional form restrictions on the counterfactual discontinuities, including constant and linear paths (Leventer and Nevo, 2024). The approach here uses the pre-treatment observables to restrict a class of post-treatment confounding paths and accounts for the statistical uncertainty of estimating the pre-treatment discontinuity in inference for the treatment effect of interest. The motivating intuition is that the observed pre-treatment confounding effect is informative of the unobserved post-treatment confounding effect.

The established theory justifies the following empirical workflow that can be easily implemented using publicly available packages in R: estimate the vector of period-specific RD jumps and its full covariance matrix using `rdrobust` (with slight modification we explain later in the paper); express the discontinuity relative to a baseline pre-treatment period; use the estimated pre-treatment path to bound the possible post-treatment deviations of the confounding discontinuity from the observed part through a researcher-specified sensitivity class; and obtain confidence sets on the partially identified treatment effect of interest in the post periods using `HonestDiD`. More broadly, this paper provides empirical researchers with practical tools for drawing transparent conclusions under assumptions that may be more

credible in empirical practice.

The remainder of the paper is organized as the following: Section 2 reviews related literature. Section 3 states the formal setup of the problem. Section 4 establishes the identification result. Section 5 discusses the estimation procedure. Section 6 proposes the inference method. Section 7 demonstrates the proposed method through simulation. Lastly, we conclude in Section 8.

2 Literature Review

This paper contributes to the econometric literature on research designs for causal inference and program evaluation that combine multiple sources of quasi-experimental variation. It is particularly related to work on regression discontinuity (RD) designs with confounding discontinuities, DiD designs under deviations from parallel trends, and partial identification and inference methods that replace exact identifying restrictions with sensitivity analysis using bounded deviations. The DiDC design lies at the intersection of these literatures.

Grembi et al. (2016) introduced the term *difference-in-discontinuities* and provided the canonical identification argument in their study of fiscal rules for Italian municipalities. Earlier and contemporaneous studies used closely related strategies in applications to study unemployment benefits, political representation, employment protection, and female political participation (Lalive, 2008; Campa, 2011; Leonardi and Pica, 2013; Casas-Arce and Saiz, 2015). The design has subsequently been used across a broad range of empirical settings. Examples include paternity leave and marital stability (Avdic and Karimi, 2018), electoral gender quotas (Bagues and Campa, 2021), gender-pay transparency (Bennedsen et al., 2022), domestic-violence legislation (Azuaga and Sampaio, 2017), the expiration of the U.S. federal assault-weapons ban (Chicoine, 2017), Norway's Reform 94 on social mobility (Bertrand et al., 2021), and algorithmic recommendations (Albright, 2024). These applications demonstrate that DiDC is empirically relevant whenever a policy threshold coincides with a pre-existing institutional or behavioral discontinuity.

The formal econometric literature on DiDC is more recent. Butts (2023) studies DiDC designs in geographic settings. The fuzzy case is studied by Millán-Quijano (2020), Larsen and Valant (2023), and Tchuente et al. (2026), who clarify the additional restrictions required when exposure to the policy of interest or the confounding policy does not change sharply at the cutoff. Takahashi (2026) studies identification of DiDC designs in cross-sectional settings, and Dong (2026) extends DiDC to distributional and quantile effects. Leventer and Nevo (2024) develop a multi-period framework for RD designs that violate the continuity assumption, for which DiDC designs can be viewed as a particular instance of such violations.

However, they do not explicitly study the DiDC design in the paper. Most directly related, Picchetti et al. (2026) provide a formal econometric analysis of the canonical sharp design with one pre-policy and one post-policy period under point identification. They additionally include partial identification results under bounded variations from time invariance and from restrictions on policy interactions using researcher specified constants. The main difference of the partial identification results from this paper is that we study the more empirically realistic case of multiple time periods, and leverage observed change in pre-treatment discontinuity to bound the post-treatment unobserved confounding effect.

This feature connects the paper to the literature on partial identification and inference under violations of conventional identifying assumptions. Bounded-variation restrictions have a long history in partial identification analysis (Manski and Pepper, 2018), and Masten and Poirier (2020) develop breakdown frontiers for reporting the strength of assumptions required to sustain a conclusion. Rambachan and Roth (2023) apply this approach to DiD and event-study designs. We adapt their framework to DiDC by treating the sequence of pre-treatment RD discontinuities as information about the counterfactual evolution of the confounding discontinuity. Additional pre-treatment periods serve not only to test the constant confounding restriction, as in Picchetti et al. (2026), but also to inform economically interpretable deviation from it and to allow analysis of how sensitive empirical conclusions are to assumption violations.

Finally, the paper contributes to the literature on nonparametric estimation and inference in RD designs. Calonico et al. (2014) develop robust bias-corrected inference for a single local-polynomial RD estimand, and subsequent work provides uniform and coverage-error refinements for such procedures (Calonico et al., 2018, 2022). Fusejima et al. (2025) establish joint asymptotic results for a vector local-polynomial RD estimators and the consistency of its full covariance matrix in the context of specification testing. We leverage their results, which are required for conditional and hybrid moment inequality tests used in Andrews et al. (2023) and Rambachan and Roth (2023), to establish inference in our setting for a fixed number of periods with a time-invariant running variable. We verify the additional conditions needed to apply their inference procedure to the nonparametric DiDC setting.

3 Setup

We observe a i.i.d. sample of n units $i = 1, \ldots, n$ over a number of time periods

$$t \in \mathcal{T} = \{-\underline{T}, \ldots, -1, 0, 1, \ldots, \bar{T}\}, \quad \underline{T} \geq 1, \quad \bar{T} \geq 1.$$

Denote $\mathcal{T}_0 = \{-\underline{T}, \ldots, 0\}$ and $\mathcal{T}_1 = \{1, ..., \bar{T}\}$ as the pre- and post-treatment periods. The total number of time periods $T = \underline{T} + \bar{T} + 1$ is fixed. Every unit has a running variable X_i with density $f(x)$ with respect to the Lebesgue measure. Denote the cutoff of interest as x_0, which we set to $x_0 = 0$ without loss of generality. Assume the running variable has support arbitrarily close to the cutoff, and limits at the cutoff are taken through the support of X_i. We consider the setup of Picchetti et al. (2026), and assume the running variable is time-invariant. For each unit i at time t, we observe its outcome, denoted as Y_{it}.

Suppose researchers are interested in the effect of a binary treatment introduced between period 0 and 1. The assignment rule of the policy of interest, denoted as D_{it}, follows standard sharp RD designs, where units with the running variable above the cutoff are assigned to treatment, and units with the running variable below the cutoff are assigned to control. Formally,

$$D_{it} = \mathbf{1}\{X_i \geq 0\}\mathbf{1}\{t \in \mathcal{T}_1\}.$$

It is worth noting that the policy of interest is only active in the post periods. In particular, all treated units are treated at the same point in time, and there is no staggered adoption.

One deviation from standard RD designs that motivates the use of DiDC designs is that there exists a confounding policy whose assignment rule is subject to the same cutoff. The confounding policy is active in every period, particularly before the policy of interest is introduced. Since the running variable is time-invariant, whether a unit is subject to the confounding policy remains unchanged over time. Formally, we denote its assignment rule as

$$Z_i = \mathbf{1}\{X_i \geq 0\}, \quad \forall t \in \mathcal{T}.$$

We define potential outcomes as a function of the contemporaneous status of both the confounding policy and the policy of interest. Let $Y_{it}(z, d)$ be the potential outcome for unit i in period t under the confounding treatment status $Z_i = z$ and $D_{it} = d$. In the post-treatment periods, the observed outcomes at each time period are related to all four potential outcomes through

$$\begin{aligned} Y_{it} =& Z_i D_{it}\, Y_{it}(1,1) + Z_i(1 - D_{it})\, Y_{it}(1,0) \\ &+ (1 - Z_i)D_{it}\, Y_{it}(0,1) + (1 - Z_i)(1 - D_{it})\, Y_{it}(0,0), \quad \forall t \in \mathcal{T}_1. \end{aligned} \tag{1}$$

In the pre-treatment period, $D_{it} = 0, \forall i, t \in \mathcal{T}_0$, Therefore, only two potential outcomes are relevant in defining the realized outcome,

$$Y_{it} = Z_i Y_{it}(1,0) + (1 - Z_i)Y_{it}(0,0), \quad \forall t \in \mathcal{T}_0. \tag{2}$$

Defining the potential outcome this way excludes anticipatory responses to future treatment status by construction. A richer dynamic potential-outcome framework could index outcomes by the entire treatment path and impose no anticipation as a separate restriction, similar to the DiD setting discussed in Roth et al. (2023). We adopt the simpler contemporaneous treatment formulation because anticipatory effects are not the focus of the analysis.

It is also important to note that in post periods, we only observe $Y_{it}(1,1)$ for units above the cutoff, and $Y_{it}(0,0)$ for units below the cutoff. The other two potential outcomes are never observed in the data due to the treatment assignment rule, but are relevant for defining the causal effect of interest below.

To fix ideas, in the classic application of DiDC designs by Grembi et al. (2016), Y_{it} is a fiscal outcome of Italian municipality i such as deficits and tax revenues, X_i is the population of each municipality with the cutoff of interest being 5000, the confounding policy is a change in the mayor's wage, and the treatment of interest is a policy on relaxation of fiscal rules introduced after 2001.

At each time period t, we observe a discontinuity at the cutoff $X_i = 0$, and we denote it as

$$\theta_t = \lim_{x \to 0^+} \mathbb{E}[Y_{it} \mid X_i = x] - \lim_{x \to 0^-} \mathbb{E}[Y_{it} \mid X_i = x]. \tag{3}$$

We also denote the causal effect of the confounding policy (in the absence of the policy of interest) in each period $t \in \mathcal{T}$ as

$$\rho_t = \mathbb{E}[Y_{it}(1,0) - Y_{it}(0,0) \mid X_i = 0]. \tag{4}$$

We are interested in identifying and doing inference on the average treatment effect of the treated under the exposure of the confounding policy in the post periods, denoted as τ_t for $t \in \mathcal{T}_1$, which we define as

$$\tau_t = \mathbb{E}[Y_{it}(1,1) - Y_{it}(1,0) \mid X_i = 0]. \tag{5}$$

One could also possibly be interested in the effects of the policy for treated units unexposed to the confounding policy. This extension requires additional assumptions, such as a no-interaction restriction, as stated in Picchetti et al. (2026, Assumption 5). Intuitively, their assumption imposes that the effect of the treatment of interest does not depend on the confounding policy. This type of extrapolation is not the focus of this paper.

4 Identification

Building upon the recent theoretical development on DiDC designs by Picchetti et al. (2026) for two time periods, we extend the identification analysis to multiple time periods.

4.1 Causal Decomposition

In the two-period design, the DiDC estimand is defined as the difference between the observed discontinuity between the pre- and post-period. We generalize this estimand by comparing the discontinuity in each period t with that in a reference period. Taking period 0 as the reference period, define the period-t DiDC estimand as

$$\beta_t = \theta_t - \theta_0, \quad \forall t \in \mathcal{T} \setminus \{0\}. \tag{6}$$

We collect the pre- and post-treatment DiDC estimands in the vectors

$$\boldsymbol{\beta}_{\text{pre}} = \begin{pmatrix} \beta_{-\underline{T}} \\ \vdots \\ \beta_{-1} \end{pmatrix}, \quad \boldsymbol{\beta}_{\text{post}} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{\bar{T}} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_{\text{pre}} \\ \boldsymbol{\beta}_{\text{post}} \end{pmatrix} \in \mathbb{R}^{\underline{T}+\bar{T}}.$$

Analogously, define the corresponding change in the confounding policy effect relative to period 0 as

$$\delta_t = \rho_t - \rho_0.$$

Collect these deviations in vectors (omitting $\delta_0 = 0$),

$$\boldsymbol{\delta}_{\text{pre}} = \begin{pmatrix} \delta_{-\underline{T}} \\ \vdots \\ \delta_{-1} \end{pmatrix}, \quad \boldsymbol{\delta}_{\text{post}} = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_{\bar{T}} \end{pmatrix}, \quad \boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\delta}_{\text{pre}} \\ \boldsymbol{\delta}_{\text{post}} \end{pmatrix} \in \mathbb{R}^{\underline{T}+\bar{T}}.$$

Finally, collect the post-treatment causal effects of the policy of interest in $\boldsymbol{\tau}_{\text{post}} = (\tau_1, ..., \tau_{\bar{T}})^\top$.

We maintain the following standard continuity condition.

Assumption 1 (Continuity)**.** All potential outcome regression functions are continuous near the cutoff. For every $t \in \mathcal{T}$ and every $(z, d) \in \{0, 1\}^2$,

$$\lim_{x \to 0^+} \mathbb{E}[Y_{it}(z, d) \mid X_i = x] = \lim_{x \to 0^-} \mathbb{E}[Y_{it}(z, d) \mid X_i = x] = \mathbb{E}[Y_{it}(z, d) \mid X_i = 0].$$

Assumption 1 adapts the standard continuity condition in Hahn et al. (2001) for iden-

tification in regression discontinuity designs to allow for the presence of the confounding treatment Z_i. Specifically, it imposes that, given the treatment status of the confounding policy and policy of interest (z, d), all other determinants of the mean outcome must evolve smoothly near the cutoff so that there are no other systematic discontinuities at each time period.

The following lemma gives the central decomposition of the multi-period DiDC estimands.

Lemma 1 (DiDC Decomposition)**.** *Under Assumption 1,*

$$\boldsymbol{\beta} = \underbrace{\begin{pmatrix} \mathbf{0}_{\underline{T}} \\ \boldsymbol{\tau}_{\text{post}} \end{pmatrix}}_{\boldsymbol{\tau}} + \underbrace{\begin{pmatrix} \boldsymbol{\delta}_{\text{pre}} \\ \boldsymbol{\delta}_{\text{post}} \end{pmatrix}}_{\boldsymbol{\delta}}. \tag{7}$$

Proof. See Appendix A.1. □

For later use, denote $\boldsymbol{\theta} = (\theta_{-\underline{T}}, ..., \theta_{\bar{T}})^\top$, and let R be the fixed $(T-1) \times T$ matrix whose row for period $t \neq 0$ is $e_t^\top - e_0^\top$, where e_t selects period t in this chronological ordering. Then

$$\boldsymbol{\beta} = R\boldsymbol{\theta}, \qquad RR^\top = I_{T-1} + \mathbf{1}_{T-1}\mathbf{1}_{T-1}^\top.$$

In particular, R has full row rank. Define also $L_{\text{post}} = \begin{pmatrix} 0_{\underline{T} \times \bar{T}} \\ I_{\bar{T}} \end{pmatrix}$, so that (7) is

$$\boldsymbol{\beta} = \boldsymbol{\delta} + L_{\text{post}}\boldsymbol{\tau}_{\text{post}}.$$

Lemma 1 decomposes the observed DiDC estimands into a causal effect component and a component capturing deviations in the effect of the confounding policy relative to the reference period. This representation is identical to the causal decomposition of event-study estimates in Rambachan and Roth (2023). In the pre-treatment periods, the policy of interest is not in place, so $\boldsymbol{\beta}_{\text{pre}} = \boldsymbol{\delta}_{\text{pre}}$. That is, the variation of the confounding policy relative to period 0 is observed as the DiDC estimand. In the post-treatment periods, $\boldsymbol{\beta}_{\text{post}} = \boldsymbol{\tau}_{\text{post}} + \boldsymbol{\delta}_{\text{post}}$. The vector $\boldsymbol{\delta}_{\text{post}}$ is the counterfactual: it describes how the effect of the confounding policy would have changed after period 0 in the absence of the policy of interest. Since the observed post-treatment discontinuities capture both this counterfactual evolution and the causal effect of the policy of interest, $\boldsymbol{\tau}_{\text{post}}$ is not identified without additional restrictions on $\boldsymbol{\delta}_{\text{post}}$.

4.2 Target Parameter and Identification

We consider the target parameter as a linear combination of the post-treatment causal effects. Let $l \in \mathbb{R}^{\bar{T}}$ be fixed and nonzero, and define the scalar target $\vartheta = l^\top \boldsymbol{\tau}_{\text{post}}$. For example, if one is interested in the treatment effect at one particular post period t, then let l be the t-th standard basis vector. Similarly, if one is interested in the average causal effect across all post-treatment periods, then $l = \left(1/\bar{T}, ..., 1/\bar{T}\right)^\top$. It is up to the researcher to specify l depending on their parameter of interest.

The conventional identifying restriction in a DiDC design discussed in both Grembi et al. (2016) and Picchetti et al. (2026) corresponds to the stability of the confounding policy effect over time. In the current notation, this requires

$$\rho_t = \rho_0, \qquad \text{for } t \in \mathcal{T}_1,$$

or, equivalently,

$$\boldsymbol{\delta}_{\text{post}} = \mathbf{0}_{\bar{T}}.$$

Under this restriction, the vector of treatment effects of interest in the post periods is point identified, $\boldsymbol{\tau}_{\text{post}} = \boldsymbol{\beta}_{\text{post}}$. This identification argument is analogous to assuming parallel trends in the DiD setting.

Similar to parallel trends, the constant confounding effect assumption is fundamentally untestable, but to provide suggestive evidence for the validity of the assumption, Picchetti et al. (2026) propose using available pre-treatment periods to estimate stacked RDs and test whether the RD estimates remain consistent across multiple periods. When they revisit the empirical application in Grembi et al. (2016), the test rejects for the specification with one of the main outcome of interest, taxes, suggesting that the confounding effects vary over time.

This issue resembles the classic pretesting problem and the violation of the parallel trend assumption in the DiD literature, summarized by Roth et al. (2023). Even if the pretest does not reject, conditioning the analysis on the result of a pretest can distort estimation and inference (Roth, 2022). More generally, the identification problem can be formulated as placing restrictions on the unobserved vector $\boldsymbol{\delta}_{\text{post}}$, potentially using the observed pre-treatment deviations $\boldsymbol{\delta}_{\text{pre}}$ to restrict the magnitude and evolution of post-treatment violations through sensitivity analysis. This is an idea that has been made popular to relax the parallel trend assumption in the DiD literature by Rambachan and Roth (2023).

Adopting the same approach, instead of assuming the constant confounding effect holds exactly, we propose assumptions that explicitly leverage the observed confounding effects in the pre-periods to bound the deviation of constant confounding, and partially identify the

treatment effect of interest in the post-periods. Picchetti et al. (2026) also discuss partial identification approaches in the spirit of Manski and Pepper (2018). However, since their theory is built on two periods, information from multiple pre-treatment time periods was not used for identification.

Formally, we follow Rambachan and Roth (2023) and assume that $\boldsymbol{\delta}$ lies in a researcher-specified set of possible deviations from constant confounding $\Delta \subseteq \mathbb{R}^{T+\bar{T}}$. The identified set for the parameter of interest ϑ is the set of values that are consistent with the observable data $\boldsymbol{\beta}$ under the restriction $\boldsymbol{\delta} \in \Delta$,

$$\mathcal{S}(\boldsymbol{\beta}, \Delta) = \left\{ l^\top \boldsymbol{\tau}_{\text{post}} \ : \exists\, \boldsymbol{\delta} \in \Delta, \boldsymbol{\tau}_{\text{post}} \in \mathbb{R}^{\bar{T}} \text{ s.t. } \boldsymbol{\beta} = \boldsymbol{\delta} + L_{\text{post}}\, \boldsymbol{\tau}_{\text{post}} \right\}. \tag{8}$$

The constant confounding effect assumption is a special case in this formulation. It is equivalent as specifying $\Delta = \left\{ \boldsymbol{\delta} : \boldsymbol{\delta}_{\text{pre}} = \boldsymbol{\beta}_{\text{pre}}, \boldsymbol{\delta}_{\text{post}} = \mathbf{0}_{\bar{T}} \right\}$. The next lemma follows directly from Rambachan and Roth (2023, Lemma 2.1).

Lemma 2. *If Δ is closed and convex, then $\mathcal{S}(\boldsymbol{\beta}, \Delta)$ is an interval, where*

$$\Big[\underbrace{l^\top \boldsymbol{\beta}_{\text{post}} - \max_{\boldsymbol{\delta} \in \Delta:\, \boldsymbol{\delta}_{\text{pre}} = \boldsymbol{\beta}_{\text{pre}}} l^\top \boldsymbol{\delta}_{\text{post}}}_{\vartheta^{lb}(\beta, \Delta)} \ , \ \underbrace{l^\top \boldsymbol{\beta}_{\text{post}} - \min_{\boldsymbol{\delta} \in \Delta:\, \boldsymbol{\delta}_{\text{pre}} = \boldsymbol{\beta}_{\text{pre}}} l^\top \boldsymbol{\delta}_{\text{post}}}_{\vartheta^{ub}(\beta, \Delta)} \Big]. \tag{9}$$

Proof. The result is immediate by rearranging terms in (8). □

4.3 Types of Restrictions

The set of possible deviations from the constant confounding effect Δ is a choice by the researcher, and it should depend on context-specific knowledge of each empirical application and the available data. Rambachan and Roth (2023) discuss a handful of polyhedral restrictions or finite unions of such restrictions, which allow for a variety of assumptions that may be relevant in empirical work. We refer to their discussion for more details. Here, we include two leading choices of Δ.

The first class of restrictions restricts the magnitude of changes in the confounding-policy effect after the policy of interest is introduced, using the changes observed in the pre-treatment periods as a benchmark. In some applications, the effect of the confounding policy may vary over time because of changes in economic conditions, implementation, or other factors that interact with the confounding policy. Researchers may nevertheless be willing to assume that these forces do not induce substantially larger changes in the post-treatment period than were observed before treatment. Rather than requiring the confounding-policy

effect to remain exactly constant, this restriction therefore uses the observed pre-treatment variation to discipline how rapidly its counterfactual post-treatment path may evolve.

Formally, for a sensitivity parameter $\bar{M} \geq 0$, define

$$\Delta^{\mathrm{RM}}(\bar{M}) = \left\{ \delta : |\delta_{t+1} - \delta_t| \leq \bar{M} \max_{s<0} |\delta_{s+1} - \delta_s| , \quad t = 0, \ldots, \bar{T} - 1 \right\}.$$

Since $\delta_{t+1} - \delta_t = \rho_{t+1} - \rho_t$, the restriction is equivalently

$$|\rho_{t+1} - \rho_t| \leq \bar{M} \max_{s<0} |\rho_{s+1} - \rho_s| , \quad t = 0, \ldots, \bar{T} - 1.$$

Hence, $\bar{M}$ controls how large a post-treatment change in the confounding-policy effect may be relative to the largest such change observed before treatment. A natural benchmark is $\bar{M} = 1$, which requires that no post-treatment change be larger than the largest pre-treatment change. Values $\bar{M} > 1$ allow the confounding-policy effect to evolve more sharply after treatment, while $\bar{M} = 0$ implies $\delta_{t+1} = \delta_t$ for every $t \geq 0$. The latter reduces to $\boldsymbol{\delta}_{\mathbf{post}} = \mathbf{0}_{\bar{T}}$, and therefore nests the conventional constant confounding effect assumption used for point identification in DiDC.

A second class of restrictions is motivated by settings in which changes in the confounding-policy effect are thought to arise from slowly evolving economic forces rather than discrete shocks. For example, the effect of a pre-existing policy at the cutoff may change gradually with local economic conditions or institutional adjustment. In such settings, requiring the confounding effect to be constant may be too restrictive, while allowing it to change arbitrarily from one period to the next may be unnecessarily weak. A natural intermediate restriction is to bound changes in the slope of the effect of the confounding policy over time.

For $M \geq 0$, define the smoothness class

$$\Delta^{\mathrm{SD}}(M) = \left\{ \delta : |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leq M, \quad t = -\underline{T} + 1, \ldots, \bar{T} - 1 \right\}.$$

Equivalently, because $\delta_t = \rho_t - \rho_0$,

$$|(\rho_{t+1} - \rho_t) - (\rho_t - \rho_{t-1})| \leq M, \quad t = -\underline{T} + 1, \ldots, \bar{T} - 1.$$

The parameter M therefore bounds the absolute change in the slope of the confounding effect from one period to the next. Smaller values of M impose a smoother counterfactual evolution of the confounding effect, whereas larger values allow greater departures from a smooth trend. In the special case $M = 0$, the confounding effect must evolve exactly linearly over time. Thus, the restriction nests linear extrapolation of the pre-treatment confounding

path as a special case.

5 Estimation

The primary object for estimation is the vector of period-specific RD estimand $\boldsymbol{\theta} = (\theta_{-\underline{T}}, ..., \theta_{\bar{T}})^\top$. The identification analysis uses differences in RD discontinuities relative to period zero, which can be obtained by applying a linear transformation, $\boldsymbol{\beta} = R\boldsymbol{\theta}$. This section discusses the estimators used in the analysis, which follow the standard robust-bias-corrected estimation procedure in the RD literature Calonico et al. (2014). We first present the specific estimator we use for bias correction, which is a higher-order implementation under the same-kernel and same-bandwidth choice. We also present the covariance estimator, following Fusejima et al. (2025), that accounts for both bias correction and dependence between outcomes observed on the same unit in different periods.

Let $W_i = (Y_{i-\underline{T}}, ..., Y_{i\bar{T}}, X_i)^\top$ be a i.i.d. sample, and let P denote its joint probability distribution. For every $t \in \mathcal{T}$, denote the conditional mean and its left and right side limits as

$$\mu_t(x) = \mathbb{E}_P[Y_{it} \mid X_i = x], \qquad \mu_{t+} = \lim_{x \to 0^+} \mu_t(x), \qquad \mu_{t-} = \lim_{x \to 0^-} \mu_t(x),$$

so that $\theta_t = \mu_{t+} - \mu_{t-}$. The expectation is taken with respect to P, and the subscript P is suppressed in the notation.

5.1 Biased-Corrected Local-Polynomial Estimator

We use polynomial expansions to approximate the conditional mean of Y_{it} separately on both sides of the cutoff using observations in a neighborhood for each period, and take the difference of the intercepts to obtain the per-period discontinuity. This approximation generally has a leading smoothing bias. Specifically, for each $t \in \mathcal{T}$ and some $\ell \geq 1$, we can estimate θ_t using the following (ℓ)th-order local polynomial estimators:

$$\begin{aligned} \widehat{\theta}_{t,\ell}(h_{t,n}) &= \widehat{\mu}_{t+,\ell}(h_{t,n}) - \widehat{\mu}_{t-,\ell}(h_{t,n}), \\ \widehat{\mu}_{t+,\ell}(h_{t,n}) &= e_0^\top \widehat{\alpha}_{t+,\ell}(h_{t,n}), \\ \widehat{\mu}_{t-,\ell}(h_{t,n}) &= e_0^\top \widehat{\alpha}_{t-,\ell}(h_{t,n}), \end{aligned} \tag{10}$$

where

$$\begin{aligned}\widehat{\alpha}_{t+,\ell}(h_{t,n}) &= \underset{\alpha\in\mathbb{R}^{\ell+1}}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbf{1}\{X_i \geq 0\}\{Y_{it} - r_\ell(X_i)^\top \alpha\}^2 \frac{K(X_i/h_{t,n})}{h_{t,n}}, \\ \widehat{\alpha}_{t-,\ell}(h_{t,n}) &= \underset{\alpha\in\mathbb{R}^{\ell+1}}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbf{1}\{X_i < 0\}\{Y_{it} - r_\ell(X_i)^\top \alpha\}^2 \frac{K(X_i/h_{t,n})}{h_{t,n}},\end{aligned} \tag{11}$$

$r_\ell(x) = (1, x, \ldots, x^\ell)^\top$ is the (ℓ)th-order polynomial expansion, $e_0 = (1, 0, ..., 0)^\top$ is the first $(\ell+1)$-dimensional unit vector. We use α for regression coefficients to distinguish them from the DiDC estimands $\boldsymbol{\beta}$. The common kernel across time periods is K, and $h_{t,n} > 0$ is the positive bandwidth sequence for period t. The bandwidth may differ across periods but is common to the two sides within a period.

It is a well-established result by Calonico et al. (2014) such that leading bias of ℓth-order polynomial estimator is of a level is of order $h_{t,n}^{\ell+1}$, and its standard deviation is of order $(nh_{t,n})^{-1/2}$. Therefore, at the MSE-optimal rate $h_{t,n} \asymp n^{-1/(2\ell+3)}$, these two terms are of the same order. Thus, the leading bias is non-negligible in the first-order distribution, and conventional inference centered at θ_t is generally invalid. Specifically, let $\mathcal{X}_n = (X_1, \ldots, X_n)^\top$. Under the standard smoothness conditions for local-polynomial estimation, the conditional bias of the degree-ℓ RD estimator satisfies

$$\mathbb{E}\Big[\widehat{\theta}_{t,\ell}(h_{t,n}) \mid \mathcal{X}_n\Big] - \theta_t = h_{t,n}^{\ell+1} B_{t,\ell}(h_{t,n}) + o_P\big(h_{t,n}^{\ell+1}\big)\,,$$

where

$$B_{t,\ell}(h_{t,n}) = \frac{\mu_{t+}^{(\ell+1)}(0)}{(\ell+1)!}\mathcal{B}_{+,\ell}(h_{t,n}) - \frac{\mu_{t-}^{(\ell+1)}(0)}{(\ell+1)!}\mathcal{B}_{-,\ell}(h_{t,n}).$$

$\mu_{t,+}^{(\ell+1)}(0)$ and $\mu_{t-}^{(\ell+1)}(0)$ denote the corresponding one-sided $(\ell+1)$th derivative at the cutoff, and $\mathcal{B}_{+,\ell}(h_{t,n})$ and $\mathcal{B}_{-,\ell}(h_{t,n})$ are the usual asymptotically bounded and observable design quantities determined by $\mathcal{X}_n$, the kernel K, and the bandwidth $h_{t,n}$.

Bias correction is important for inference. The general robust bias-correction construction estimates $B_{t,\ell,P}(h_{t,n})$ with a degree-$(\ell+1)$ regression and a pilot bandwidth $b_{t,n}$, subtracts the estimated leading bias, and computes the variance of the entire corrected estimator. If $h_{t,n}/b_{t,n} \to c_t \in (0, \infty)$, this variance includes the variance of the estimated bias and its covariance with the original estimator.

We leverage an equivalence result for bias correction discussed in Calonico et al. (2014, Remark 7). This implementation sets $b_{t,n} = h_{t,n}$, and uses the same kernel in the original and bias estimation regressions. Under this specification, the bias-corrected degree-ℓ estimator is numerically equivalent to the uncorrected degree-$(\ell+1)$ local-polynomial estimator using

the same MSE-optimal order bandwidth for degree ℓ, denoted as

$$\widehat{\theta}_{t,\ell+1}(h_{t,n}) = \widehat{\theta}_{t,\ell}(h_{t,n}) - h_{t,n}^{\ell+1}\widehat{B}_{t,\ell}(h_{t,n}, b_{t,n}), \tag{12}$$

where $\widehat{\theta}_{t,\ell+1}(h_{t,n})$ is defined analogously as in Equation (10)-(11). For example, when $\ell = 1$, the bandwidth is selected to be MSE-optimal for the local-linear RD estimator, but the reported point estimate and standard error are based on a local-quadratic fit evaluated at that bandwidth.

This approach of increasing the polynomial order for bias correction is the same-bandwidth and same-kernel special case of robust bias correction in Calonico et al. (2014). The difference between the bandwidth-selection order and the estimation order is essential. The intuition of why this works is that if $h_{t,n} \asymp n^{-1/(2\ell+3)}$, then the leading bias of the order-ℓ estimator is first-order non-negligible. However, the remaining bias term from the order-$(\ell+1)$ fit is of order $h_{t,n}^{\ell+2}$, and $\sqrt{nh_{t,n}}\, h_{t,n}^{\ell+2} \to 0$. Hence, the leading bias vanishes.

Stacking the per-period estimator, we denote the estimator we use for the rest of the paper as

$$\widehat{\boldsymbol{\theta}} = \left(\widehat{\theta}_{t,\ell+1}(h_{t,n}) : t \in \mathcal{T}\right)^{\top}. \tag{13}$$

As a result, we can obtain an estimator for $\boldsymbol{\beta}$ by applying the linear transformation, denoted as

$$\widehat{\boldsymbol{\beta}} = R\widehat{\boldsymbol{\theta}}.$$

5.2 Covariance Estimator

The inference procedure in the following section requires the full sampling covariance matrix, rather than only the period-specific standard errors. We use the nearest-neighbor covariance estimator of Calonico et al. (2014), which may perform better than plugging in the corresponding residuals of nonparametric regressions in finite samples. The covariance estimator uses the higher-order fit, accounting for bias correction.

We start by denoting J as the number of neighbors, and let $N_J^+(i)$ and $N_J^-(i)$ denote the J closest other units to unit i in the running variable, restricted to the same side of the cutoff. For period t and u, define the estimator of the conditional variance between Y_{it} and

Y_{iu} at each side of the cutoff as

$$\widehat{\sigma}^2_{tu+}(X_i) = \mathbf{1}\{X_i \geq 0\}\frac{J}{J+1}\left(Y_{it} - \frac{1}{J}\sum_{j\in N_J^+(i)} Y_{jt}\right)\left(Y_{iu} - \frac{1}{J}\sum_{j\in N_J^+(i)} Y_{ju}\right),$$
$$\widehat{\sigma}^2_{tu-}(X_i) = \mathbf{1}\{X_i < 0\}\frac{J}{J+1}\left(Y_{it} - \frac{1}{J}\sum_{j\in N_J^-(i)} Y_{jt}\right)\left(Y_{iu} - \frac{1}{J}\sum_{j\in N_J^-(i)} Y_{ju}\right).$$

The factor $J/(J+1)$ corrects the additional variance introduced by subtracting the average of J neighboring observations. We also define

$$\begin{aligned}
\Gamma_{+,\ell+1}(h) &= \frac{1}{n}\sum_i \mathbf{1}\{X_i \geq 0\}K_h(X_i)r_{\ell+1}(X_i/h)r_{\ell+1}(X_i/h)^\top, \\
\Gamma_{-,\ell+1}(h) &= \frac{1}{n}\sum_i \mathbf{1}\{X_i < 0\}K_h(X_i)r_{\ell+1}(X_i/h)r_{\ell+1}(X_i/h)^\top, \\
\widehat{\Psi}_{tu+,\ell+1}(h_t,h_u) &= \frac{1}{n}\sum_i \mathbf{1}\{X_i \geq 0\}K_{h_t}(X_i)K_{h_u}(X_i)r_{\ell+1}(X_i/h_t)r_{\ell+1}(X_i/h_u)^\top\widehat{\sigma}_{tu+}(X_i), \\
\widehat{\Psi}_{tu-,\ell+1}(h_t,h_u) &= \frac{1}{n}\sum_i \mathbf{1}\{X_i < 0\}K_{h_t}(X_i)K_{h_u}(X_i)r_{\ell+1}(X_i/h_t)r_{\ell+1}(X_i/h_u)^\top\widehat{\sigma}_{tu-}(X_i),
\end{aligned} \tag{14}$$

where $K_h(x) = K(x/h)/h$. Then, the estimated sampling covariance of the period-specific discontinuities can be defined as the $T \times T$ matrix

$$\widehat{V}_n \equiv \widehat{\mathrm{Var}}\left(\widehat{\boldsymbol{\theta}}\right) = \left\{\widehat{\mathrm{Cov}}_{tu+,\ell+1} + \widehat{\mathrm{Cov}}_{tu-,\ell+1}\right\}_{t,u\in\mathcal{T}}, \tag{15}$$

where

$$\begin{aligned}
\widehat{\mathrm{Cov}}_{tu+,\ell+1} &= \frac{1}{n}e_0^\top\Gamma_{+,\ell+1}(h_{t,n})^{-1}\widehat{\Psi}_{tu+,\ell+1}(h_{t,n},h_{u,n})\Gamma_{+,\ell+1}(h_{u,n})^{-1}e_0, \\
\widehat{\mathrm{Cov}}_{tu-,\ell+1} &= \frac{1}{n}e_0^\top\Gamma_{-,\ell+1}(h_{t,n})^{-1}\widehat{\Psi}_{tu-,\ell+1}(h_{t,n},h_{u,n})\Gamma_{-,\ell+1}(h_{u,n})^{-1}e_0.
\end{aligned} \tag{16}$$

This matrix is symmetric and positive semidefinite. Because the observations used on the two sides of the cutoff are disjoint, their centered regression errors have zero conditional cross-covariance. The sandwich matrices in (16) use the degree-$(\ell+1)$ regressions and therefore already incorporate the additional variability induced by bias estimation. The off-diagonal entries of $\widehat{V}_n$ capture the serial covariance between Y_{it} and Y_{iu} for the same unit. This estimator mimics the covariance estimator in Fusejima et al. (2025, Appendix A.4). The bandwidth factors are already included in $K_{h_t}K_{h_u}$ in (14), and the factor $1/n$ in (16) puts the result on the sampling scale.

For the vector of DiDC estimator, $\widehat{\boldsymbol{\beta}} = R\widehat{\boldsymbol{\theta}}$, define its covariance as

$$\widehat{\Omega}_n = R\widehat{V}_n R^\top. \tag{17}$$

The off-diagonal entries preserve dependence across periods and the uncertainty from subtracting the common baseline. $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Omega}_n$ are the point estimates and sampling covariance used in inference.

5.3 Asymptotic Properties

We then establish the asymptotic properties of the point estimates of the vector $\widehat{\boldsymbol{\theta}}$ and its variance. We start by imposing the necessary assumptions.

Assumption 2 (Regularity Conditions)**.** For some $\kappa_0 > 0$, the following holds true for $(-\kappa_0, 0) \cup (0, \kappa_0)$ around the cutoff $\bar{x} = 0$:

(i) The density $f(x)$ is continuous and bounded away from zero in the neighborhood $(-\kappa_0, \kappa_0)$.

(ii) For $t \in \mathcal{T}$, $\mathbb{E}[|Y_{it}|^4 \mid X_i = x]$ is bounded, and $\mu_t(x)$ is S times continuously differentiable on each closed one-sided neighborhood.

(iii) For $t, u \in \mathcal{T}$, $\mathrm{Var}(Y_{it} \mid X_i = x)$ and $\mathrm{Cov}(Y_{it}, Y_{iu} \mid X_i = x)$ are Lipschitz continuous on each closed one-sided neighborhood. Additionally, each $\mathrm{Var}(Y_{it} \mid X_i = x)$ is bounded away from zero.

These are fairly standard regularity assumptions for the data generating process in regression discontinuity designs. Assumption 2(i) ensures that the density is continuous near the cutoff, and ensures the existence of observations arbitrarily close to the cutoff in large samples. Assumption 2(ii) and (iii) ensure that the conditional expectations, variances, and covariances are well defined, and have sufficient smoothness with the existence of fourth moments. We do not restrict the limits of these quantities from the left and right sides to be the same. The Lipschitz continuity assumption is needed to ensure consistency of the nearest-neighbor covariance estimator. Lastly, we require only the variances to be bounded away from zero. Cross-period covariances can be zero or negative.

Assumption 3 (Kernel)**.** The kernel K is symmetric, bounded, nonnegative, supported on $[-1, 1]$, and positive and continuous on $(-1, 1)$.

This is also a set of standard assumptions on the kernels commonly used in empirical work. Both the triangular and uniform kernels satisfy it.

Assumption 4 (Bandwidths Rates)**.** Given fixed polynomial order $\ell \geq 1$ and neighbor count $J \geq 1$, for positive deterministic bandwidths, denote $h_{\min,n} = \min_t h_{t,n}$ and $h_{\max,n} = \max_t h_{t,n}$. We assume for $n \to \infty$

$$h_{\max,n} \to 0, \qquad nh_{\min,n} \to \infty, \qquad nh_{\max,n}^{2\ell+5} \to 0, \qquad \frac{h_{t,n}}{h_{u,n}} \to c_{tu} \in (0,\infty) \quad (t,u \in \mathcal{T}).$$

The first three requirements are the same rate conditions for the local polynomial mean-estimator rate conditions in (Fusejima et al., 2025, Theorem 3.1). Intuitively, we use a shrinking window, but the window cannot shrink too fast such that we ensure even the estimator using the narrowest window must still have enough observations near the cutoff. The window also cannot shrink too slowly such that the smoothing bias must vanish faster than the standard error. We additionally require all the pairwise bandwidth ratios to converge to positive and finite constants. This is stronger than the conditions in Fusejima et al. (2025) in the sense that it imposes that the estimators at each period converges at the same rate, which we require for the subsequent inference result. However, it does not require equal bandwidths across periods. Every period can have its own bandwidth, chosen to reflect its outcome's curvature and variability, but it needs to converge at the same rate.

Proposition 1 (Joint Normality and Covariance Consistency)**.** *Suppose that Assumptions 2–4 with $S \geq \ell+3$, then for the estimator using the higher-order $(\ell+1)$-polynomial implementation,*

$$\sqrt{n}D_n(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}_T(0, V_\theta), \qquad nD_n\widehat{V}_nD_n \xrightarrow{p} V_\theta, \tag{18}$$

where $D_n = \operatorname{diag}\{\sqrt{h_{t,n}} : t \in \mathcal{T}\}$ is the period-specific normalization, and V_θ is the asymptotic variance. Its exact form is specified in Appendix A.2.

Proof. See Appendix A.2. This result is an application of Remark 7 in Calonico et al. (2014) and Theorem A.1 in Fusejima et al. (2025). □

The quantity used in inference, $\widehat{\boldsymbol{\beta}}$, subtracts the estimated period-zero discontinuity from all other periods. The following corollary states the asymptotic properties of its limiting distribution and the covariance estimator.

Corollary 1. *Under the conditions of Proposition 1, let $D_{\beta,n} = \operatorname{diag}\{\sqrt{h_{t,n}} : t \neq 0\}$. We can obtain*

$$\sqrt{n}D_{\beta,n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}_{T-1}(0, V_\beta), \qquad nD_{\beta,n}\widehat{\Omega}_nD_{\beta,n} \xrightarrow{p} V_\beta. \tag{19}$$

where $V_\beta = R_*V_\theta R_*^\top$. R_* is defined as the limit of R_n where $R_n = D_{\beta,n}RD_n^{-1}$. The row of R_* for $t \neq 0$ is $e_t^\top - \sqrt{c_{t0}}e_0^\top$ and e_t is the standard basis vector that selects period t.

Proof. See Appendix A.3. □

Proposition 1 together with Corollary 1 establishes the asymptotic properties of the DiDC estimators for multiple time periods and their covariances that we use for inference.

Remark 1. The results so far use deterministic bandwidths satisfying the rate condition in Assumption 4. In particular, we allow different bandwidths for each period. Following Fusejima et al. (2025) for implementation, we use the MSE-optimal bandwidth selector in Calonico et al. (2014) for the degree-(ℓ) RD estimator and estimate each discontinuity using a degree-($\ell+1$) local polynomial at that bandwidth. For example, a common choice in empirical work is to use the MSE-optimal bandwidth local-linear regressions, and report the RD estimates as differences between local-quadratic intercepts. The same selected bandwidths enter the full covariance estimator and are held fixed when conducting sensitivity analysis. Under the selector conditions in Calonico et al. (2014, Supplement, Theorem A.4), applied period by period, the selected bandwidths are asymptotically equivalent to deterministic sequences of order $n^{-1/(2\ell+3)}$, with positive, finite period-specific constants. These deterministic sequences satisfy the rate conditions in Assumption 4.

6 Inference

Sensitivity restrictions discussed above describe which changes in the effect of the confounding are economically plausible. Inference must also account for sampling error in estimating the discontinuities used in those restrictions. We construct confidence sets for $\vartheta = l^\top \boldsymbol{\tau}_{\text{post}}$, the partially identified linear combination of treatment effects at the cutoff under exposure to the confounding policy in the post periods by inverting moment inequality tests. The construction follows that of Rambachan and Roth (2023), which uses the conditional and conditional-least-favorable hybrid moment inequality ARP tests in Andrews et al. (2023). This paper offers another application of these moment inequality tests, extending their use to sensitivity analysis in DiDC designs. The confidence sets can be obtained by supplying the point estimates $\widehat{\boldsymbol{\beta}}$ and full sampling covariance $\widehat{\Omega}_n$ directly to the `HonestDiD` package, which is easy to implement.

We briefly discuss the intuition of the tests here and refer the reader to Rambachan and Roth (2023) and Andrews et al. (2023) for full details. Firstly, consider a nonempty restriction of the form $\Delta = \{\boldsymbol{\delta} : A\boldsymbol{\delta} \leq d\}$, with fixed $A \in \mathbb{R}^{m\times(T-1)}$, $d \in \mathbb{R}^m$, and nonzero rows of A. This form is regarded as a class of polyhedral restrictions. The smoothness class is polyhedral, and the relative-magnitude class is a finite union of polyhedra.

Recall that $\boldsymbol{\beta} = \boldsymbol{\delta} + L_{\text{post}}\boldsymbol{\tau}_{\text{post}}$, where $\beta_t = \theta_t - \theta_0$ is the change in the observed period-specific discontinuity and $\delta_t = \rho_t - \rho_0$ is the change in the confounding-policy effect relative to the baseline period 0. Consider testing the null hypothesis at a candidate $H_0 : \vartheta = v$ under the restriction of $\boldsymbol{\delta} \in \Delta$. This test asks whether there exists some post-treatment effect vector with linear combination v that is consistent with the observable data and the restriction on the path of the confounding effect. To express this as a moment inequalities test, we follow Rambachan and Roth (2023, Section 3.1). First, suppose that the normal approximation holds exactly: $\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \Omega_n)$, with known covariance matrix Ω_n. Define $Y_n = A\widehat{\boldsymbol{\beta}} - d$. Then

$$\mathbb{E}_{\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \Omega_n)} \left[Y_n - AL_{\text{post}}\boldsymbol{\tau}_{\text{post}} \right] = A\boldsymbol{\beta} - d - AL_{\text{post}}\boldsymbol{\tau}_{\text{post}} = A\boldsymbol{\delta} - d.$$

Thus, $\boldsymbol{\delta} \in \Delta$ is equivalent to $\mathbb{E}_{\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \Omega_n)} \left[Y_n - AL_{\text{post}}\boldsymbol{\tau}_{\text{post}} \right] \leq 0$. The null therefore requires these moment inequalities to hold for some $\boldsymbol{\tau}_{\text{post}}$ satisfying $l^\top \boldsymbol{\tau}_{\text{post}} = v$,

$$H_0(v): \quad \exists\, \boldsymbol{\tau}_{\text{post}} \in \mathbb{R}^{\bar{T}} \quad \text{s.t.} \quad \begin{cases} l^\top \boldsymbol{\tau}_{\text{post}} = v, \\ \mathbb{E}_{\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \Omega_n)} [Y_n - AL_{\text{post}}\boldsymbol{\tau}_{\text{post}}] \leq 0. \end{cases} \tag{20}$$

At the population level, the null is equivalent to $v \in \mathcal{S}(\boldsymbol{\beta}, \Delta)$. The actual estimator need not be exactly normally distributed or exactly unbiased in finite samples. The asymptotic coverage result below requires a centered normal approximation and consistent estimation of the covariance matrix.

With one post-treatment period and $l = 1$, testing $\vartheta = v$ fixes the treatment effect at v and asks whether the remaining change in the RD jump can be explained by an allowed change in the confounding effect. With multiple post-treatment periods, the restriction $l^\top \boldsymbol{\tau}_{\text{post}} = v$ leaves $\bar{T} - 1$ nuisance parameters. For example, with two post-treatment periods and $l = (1/2, 1/2)^\top$, fixing the average effect leaves one period effect free, with the other determined by the average. The test asks whether the moment inequalities hold for some value of these nuisance parameters, accounting for sampling uncertainty. The test rejects only if even the best-fitting allowed explanation with average effect v leaves a discrepancy too large to plausibly reflect sampling error. If it does not reject, that means v remains a possible explanation. Other average effects may also remain possible, and therefore the confidence set collects all the values the test does not reject.

Building on this intuition, the test statistic is the smallest attainable value of the largest standardized inequality violation, after choosing the nuisance effects most favorably to the candidate value. A large positive value means that even this choice leaves substantial dis-

agreement. The conditional test identifies the combination of inequalities providing the strongest evidence and adjusts its critical value for selecting that combination using the same data. This is what makes this test a conditional test. When several combinations provide similarly strong evidence, conditioning on which is strongest can discard useful information. The hybrid adds a first-stage least-favorable test, calibrated with the relevant mean inequalities set to equality, that can reject a sufficiently large overall discrepancy. If the first stage does not reject, an adjusted conditional test follows. For fixed $0 < \kappa < \alpha$, the first stage has level κ and the second has conditional level $(\alpha - \kappa)/(1 - \kappa)$, additionally conditioning on first-stage acceptance. For implementation, in the `HonestDiD` R package, the underlying confidence set routine of the hybrid test is selected by default to $\kappa = \alpha/10$. The researcher supplies the overall level α rather than separately calibrating the two tests.

The setup of our testing problem is analogous to that of Rambachan and Roth (2023) in the DiD setting. We refer readers to their Sections 3.1–3.2 for the reparameterization of the null in terms of unrestricted linear nuisance parameters, the linear-programming formulation of the test statistic, and the conditional and hybrid critical value calculations. Both procedures use $\widehat{\boldsymbol{\beta}}$ and its full estimated sampling covariance $\widehat{\Omega}_n$. We denote the tests as $\psi^{C}_{\alpha,n}(v)$ and $\psi^{C\text{-}LF}_{\kappa,\alpha,n}(v)$, with one indicating rejection. Their confidence sets are

$$C^{C}_{\alpha,n} = \{v : \psi^{C}_{\alpha,n}(v) = 0\}, \qquad C^{C\text{-}LF}_{\kappa,\alpha,n} = \{v : \psi^{C\text{-}LF}_{\kappa,\alpha,n}(v) = 0\}. \tag{21}$$

Additionally, we maintain a regularity condition on the geometry of the inequalities as in Rambachan and Roth (2023, Assumption 5).

Assumption 5 (Regularity Conditions on A)**.** For nonnegative integers $k_1 + k_2 = T - 1$, the matrix A can be written as $\mathsf{T}Q$, where Q has full row rank and $\mathsf{T} = \begin{pmatrix} I_{k_1} & 0 \\ -I_{k_1} & 0 \\ 0 & I_{k_2} \end{pmatrix}$. Zero-dimensional blocks are omitted. The symbol T here is a matrix, distinct from the number of periods T.

The assumption requires that any linear dependence among the rows of A arises only from pairs of inequalities with opposite signs. This structure accommodates many restrictions of interest, including $\Delta^{SD}(M)$ and the polyhedra defining $\Delta^{RM}(M)$.

Proposition 2 (Pointwise Size Control)**.** *For each data-generating process P compatible with the causal decomposition and $\boldsymbol{\delta} \in \Delta$, fix $0 < \kappa < \alpha < 1/2$. Under Assumptions 2-5, and with V_θ nonsingular, the sets in (21) satisfy*

$$\begin{aligned} \liminf_{n\to\infty} \inf_{v \in \mathcal{S}(\boldsymbol{\beta},\Delta)} P\{v \in C^{C}_{\alpha,n}\} \geq 1 - \alpha, \\ \liminf_{n\to\infty} \inf_{v \in \mathcal{S}(\boldsymbol{\beta},\Delta)} P\{v \in C^{C\text{-}LF}_{\kappa,\alpha,n}\} \geq 1 - \alpha. \end{aligned} \tag{22}$$

Proof. See Appendix A.4. □

The guarantee is pointwise in the data-generating process and uniform over candidate scalar effects in its identified set. Fixing a data-generating process, whichever admissible confounding path generates the observed discontinuities, the corresponding treatment effect has asymptotic coverage of at least $1-\alpha$. This proposition justifies the use of the conditional and hybrid tests with coverage guarantees in the context of doing sensitivity analysis in DiDC designs. It allows convenient implementation using the familiar RD estimation procedure together with existing packages for sensitivity analysis.

7 Simulation

We conduct Monte Carlo simulations to demonstrate the performance of the procedure in practice. We consider $n = 1000$ independent units observed in four periods, $t \in \{-2, -1, 0, 1\}$, with treatment introduced between periods 0 and 1. Following the simulation design in Picchetti et al. (2026), we draw a time-invariant running variable from a rescaled beta distribution and generate outcomes using side-specific fifth-order polynomials and Gaussian errors. We allow the conditional mean functions to change over time and errors to be correlated within units. Appendix B provides the complete data generating processes and implementation details.

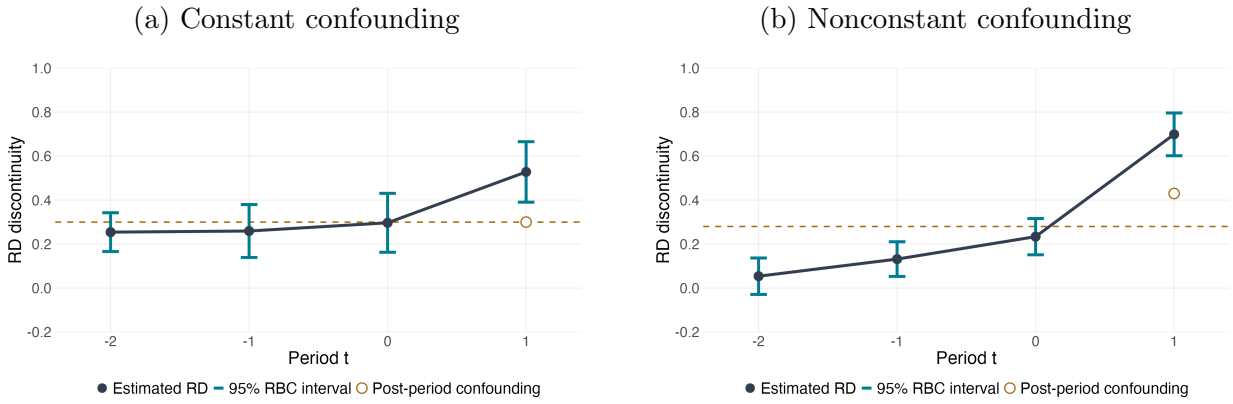
We consider the exercise of conducting inference for the treatment effect at period 1. Both DGPs have a true treatment effect $\vartheta = \tau_1 = 0.20$. In the constant-confounding design, $\rho_t = 0.30$ in every period. The true RD discontinuities at each time period are $\boldsymbol{\theta} = (0.30, 0.30, 0.30, 0.50)$. In the nonconstant confounding design, the confounding effects are $(0.10, 0.18, 0.28, 0.43)$. The corresponding true discontinuities $\boldsymbol{\theta}$ are $(0.10, 0.18, 0.28, 0.63)$. Thus, the change in the discontinuity between periods 0 and 1 equals 0.20 in the first design and 0.35 in the second; the latter combines the treatment effect with a 0.15 increase in the confounding effect.

For implementation, we follow the recommendation of Rambachan and Roth (2023) to use the hybrid test for generic restrictions of the sensitivity class. We compare nominal 95% conditional-least-favorable (C-LF) confidence sets under the relative magnitude (RM) and smoothness (SD) restrictions with the confidence interval of Picchetti et al. (2026). Their DiDC estimator uses two periods, one before and one after treatment, and identifies the treatment effect under a constant confounding effect at the cutoff. With a time-invariant running variable, they estimate the discontinuity in the within-unit outcome difference, $Y_{i1} - Y_{i0}$, using local polynomial regression and robust bias-corrected inference. We implement

their confidence interval only using data from periods 0 and 1. We refer readers to their paper for the details of their method.

Figure 1 presents the raw period-specific RD 95% confidence intervals for one simulated sample of each DGP. The solid blue dots are the estimated discontinuities at each period, and the hollow yellow dot is the true confounding discontinuity in the post period. The pre-treatment discontinuities suggest a constant or increasing confounding path. When the true confounding effect is not constant over time, the standard two-period DiDC design would falsely assume the unobserved confounding effect in period 1 is at the level of the dashed line, which induces a bias captured by the vertical difference between the dashed line and the hollow yellow dot.

Figure 1: Estimated Discontinuities in Each Period



Notes: Each panel uses one simulated sample of 1000 units. Vertical bars are individual 95% robust bias-corrected RD confidence intervals. Filled dots, connected by solid lines, mark the estimated discontinuity in each period; the hollow dot marks the true post-treatment confounding effect, ρ_1. The horizontal dashed line marks ρ_0, the post-treatment confounding level implied by the constant-confounding assumption $\rho_1 = \rho_0$.

For the same simulated sample, Figure 2 compares our nominal 95% confidence sets with the 95% confidence interval proposed by Picchetti et al. (2026). The confidence sets from our sensitivity analysis approach are generally wider than the two-period benchmark because the sensitivity analysis also accommodates uncertainty about estimating the pre-period confounding effects. Under constant confounding, the population identified set under RM is the singleton $\{0.20\}$ for every finite M, because $\max\{|\rho_{-1} - \rho_{-2}|, |\rho_0 - \rho_{-1}|\} = 0$. However, the sample confidence sets still widen as M increases because pre-period discontinuities are estimated with error, so the maximum observed deviation from constant confounding might not be zero in finite samples. Under the SD restriction, $M = 0$ imposes a linear confounding effect with an unknown slope, which must be estimated even when its true slope is zero. In this constant confounding sample, the SD confidence set at $M = 0$ is approximately $[-0.008, 0.369]$, whereas the two-period benchmark is $[0.108, 0.278]$. Every displayed SD

set contains zero, so this sample does not establish a positive treatment effect under the maintained SD restrictions.

In the nonconstant confounding DGP, the two-period DiDC estimand is 0.35, whereas the true treatment effect is 0.20. In the displayed sample, Picchetti et al. (2026)'s confidence interval misses the true treatment effect. The population RM identified set also excludes 0.20 when $M < 1.5$. Similarly, the displayed RM confidence sets at $M = 0, 0.5$, and 1 miss the true effect, although such exclusion need not occur in every sample. Under the SD restriction, the population identified set contains the true effect when $M \geq 0.05$. The pre-treatment confounding deviations suggest an approximately linear trend, and SD restricts how much the subsequent change can depart from the preceding change. Relative to the true last pre-treatment change of 0.10, the confidence set at $M = 0.15$, which allows the post-treatment change to differ by as much as 150%, still remains strictly above zero. Thus, the conclusion of the treatment effect being positive survives substantial departures from linear extrapolation in this sample. SD is particularly informative for this design because it restricts changes in the slope of confounding rather than its level or the magnitude of its first differences.

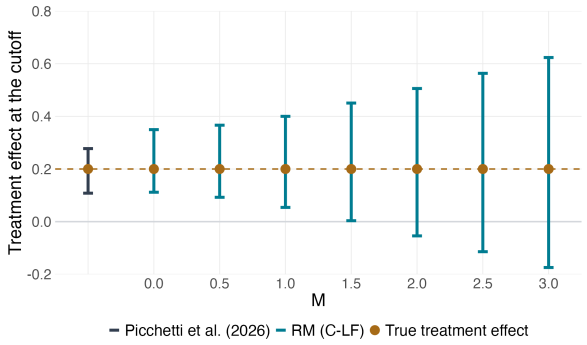
Containing the true effect in a single sample does not establish proper coverage. Figure 3 reports empirical coverage of $\vartheta = 0.20$ across 2,000 replications for each DGP. The dashed line shows coverage of the confidence interval proposed by Picchetti et al. (2026), and the dots show coverage of our confidence sets at each sensitivity value. Under constant confounding, Picchetti et al. (2026)'s interval achieves 94.05% coverage, close to the nominal 95% level. Similarly, our confidence sets achieve coverage close to or above 95% at every displayed sensitivity value. This observation is consistent with correct specification, as the constant confounding path satisfies both the identifying assumption in the canonical DiDC designs and the RM and SD restrictions for every nonnegative sensitivity bound.

Under nonconstant confounding, Picchetti et al. (2026)'s interval severely undercovers, with coverage falling to 7.30%. RM confidence sets also undercover at small $\bar{M}$, although their coverage exceeds that of the two-period benchmark. Coverage rises from 32.00% at $\bar{M} = 0$ to 99.15% at $\bar{M} = 1.5$. Coverage under the SD restriction is close to or above 95% for all sensitivity parameters in this analysis. For $M \geq 1.5$ under the RM restriction and $M \geq 0.05$ under the SD restriction, the assumptions are valid in the sense that the population identified set contains the true treatment effect. At every displayed sensitivity value satisfying these thresholds, empirical coverage achieves the nominal 95% level.

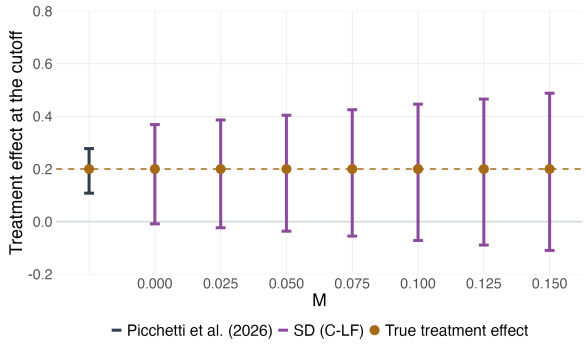
These results demonstrate how pre-treatment information can guide sensitivity analysis when constant confounding is difficult to justify. Imposing restrictions that accommodate the evolving confounding path can yield informative confidence sets and have better coverage.

Figure 2: Comparison Confidence Sets for One Realized Sample

Constant confounding

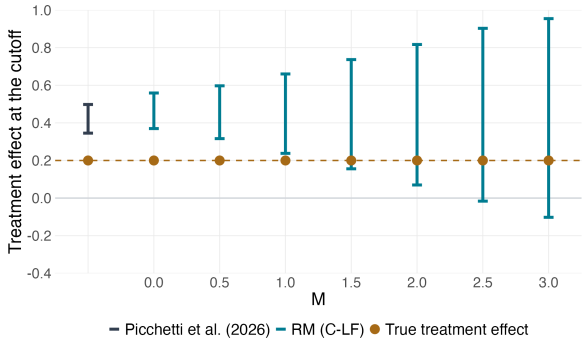


(a) $\Delta^{\mathrm{RM}}(M)$

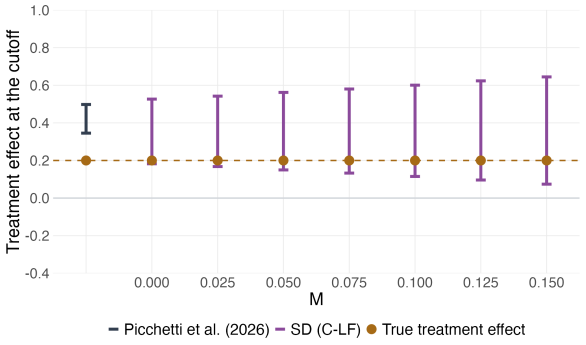


(b) $\Delta^{\mathrm{SD}}(M)$

Nonconstant confounding



(c) $\Delta^{\mathrm{RM}}(M)$



(d) $\Delta^{\mathrm{SD}}(M)$

Notes: The intervals are nominal 95% intervals/sets computed using the same sample realizations as in Figure 1. Yellow dots indicate the true value of the parameter of interest $\vartheta = 0.20$. RM and SD sensitivity parameters bound different features of the confounding path and have different scales.

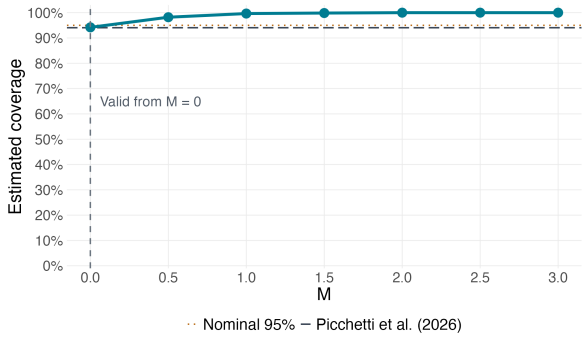
In particular, the SD confidence sets in this example remain strictly positive even when the allowed deviations from linear extrapolation are substantial. The example illustrates how restrictions informed by pre-treatment evidence and the empirical context can support credible causal conclusions and reliable inference under more plausible assumptions about confounding.

8 Conclusion

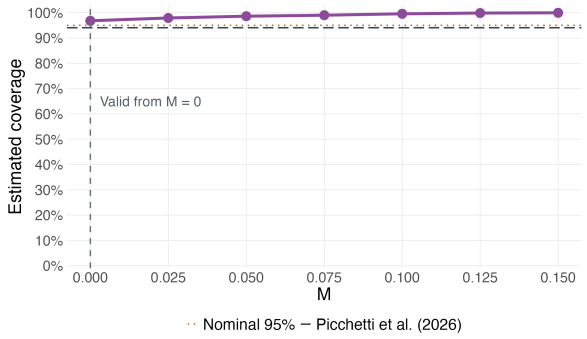
This paper develops a framework for sensitivity analysis and provide a inference procedure in difference-in-discontinuities (DiDC) designs with multiple time periods. Conventional DiDC identification relies on the assumption that the discontinuity induced by a pre-existing

Figure 3: Comparison of Finite Sample Coverage Probability

Constant confounding

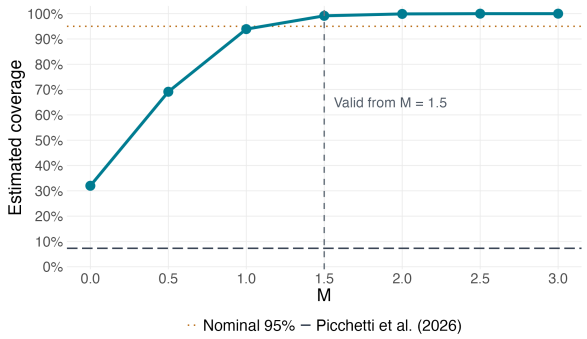


(a) $\Delta^{\mathrm{RM}}(M)$

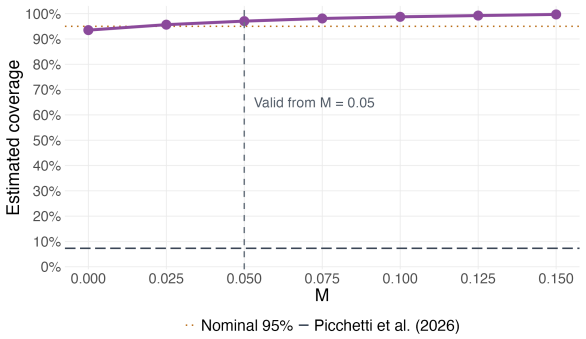


(b) $\Delta^{\mathrm{SD}}(M)$

Nonconstant confounding



(c) $\Delta^{\mathrm{RM}}(M)$



(d) $\Delta^{\mathrm{SD}}(M)$

Notes: Each dot shows the coverage of $\vartheta = 0.20$ across 2000 replications with $n = 1000$. Horizontal reference lines show nominal 95% coverage and Picchetti et al. (2026) finite-sample coverage. "Valid from" marks the smallest sensitivity value M for which the population identified set contains the true effect.

confounding policy remain constant over time in the absence of the treatment of interest, which can be hard to defend in practice. We relax this assumption by allowing the confounding discontinuity to evolve over time, subject to restrictions that bound its deviations from being constant. By exploiting information from multiple pre-treatment periods, our framework characterizes the identified sets for treatment effects under alternative sensitivity restrictions, making explicit how causal conclusions depend on the strength of the identifying assumptions. Leveraging the joint asymptotic theory for local-polynomial RD estimates, we apply the conditional and hybrid moment inequality tests in the DiDC setting. Under the relevant regularity conditions, this allows researchers to construct confidence sets that account for both sampling uncertainty and uncertainty arising from deviations from the conventional identifying assumption. Importantly, the proposed approach builds on existing

estimation and inference methods, providing a practical workflow that can be implemented using familiar regression discontinuity estimation and sensitivity analysis packages.

More broadly, our framework provides a way to assess the credibility of causal conclusions in DiDC designs without requiring the conventional identifying assumption to hold exactly. Rather than treating identification as an all-or-nothing question, researchers can examine how their conclusions change under different assumptions about the evolution of confounding discontinuities. While the inference results established here provide pointwise-in-DGP asymptotic size control, extending these results to uniform inference and settings with time-varying running variables remains directions for future research.

References

Albright, A. (2024). The hidden effects of algorithmic recommendations. Working Paper 104, Federal Reserve Bank of Minneapolis, Opportunity and Inclusive Growth Institute.

Andrews, I., Roth, J., and Pakes, A. (2023). Inference for linear conditional moment inequalities. *Review of Economic Studies*, 90(6):2763–2791.

Avdic, D. and Karimi, A. (2018). Modern family? paternity leave and marital stability. *American Economic Journal: Applied Economics*, 10(4):283–307.

Azuaga, F. L. and Sampaio, B. R. (2017). Violência contra mulher: O impacto da lei maria da penha sobre o feminicídio no brasil. In *Anais do 45º Encontro Nacional de Economia*. Associação Nacional dos Centros de Pós-Graduação em Economia.

Bagues, M. F. and Campa, P. (2021). Can gender quotas in candidate lists empower women? evidence from a regression discontinuity design. *Journal of Public Economics*, 194:104315.

Bennedsen, M., Simintzi, E., Tsoutsoura, M., and Wolfenzon, D. (2022). Do firms respond to gender pay gap transparency? *Journal of Finance*, 77(4):2051–2091.

Bertrand, M., Mogstad, M., and Mountjoy, J. (2021). Improving educational pathways to social mobility: Evidence from Norway's Reform 94. *Journal of Labor Economics*, 39(4):965–1010.

Butts, K. (2023). Geographic difference-in-discontinuities. *Applied Economics Letters*, 30(5):615–619.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2022). Coverage error optimal confidence intervals for local polynomial regression. *Bernoulli*, 28(4):2998–3022.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Campa, P. (2011). Gender quotas, female politicians and public expenditures: Quasi-experimental evidence. *SSRN Electronic Journal*.

Casas-Arce, P. and Saiz, A. (2015). Women and power: Unpopular, unwilling, or held back? *Journal of Political Economy*, 123(3):641–669.

Chicoine, L. E. (2017). Homicides in Mexico and the expiration of the U.S. federal assault weapons ban: A difference-in-discontinuities approach. *Journal of Economic Geography*, 17(4):825–856.

Dong, Y. (2026). Quantile treatment effects in difference-in-discontinuities designs. *The Econometrics Journal*, page utag028.

Eggers, A. C., Freier, R., Grembi, V., and Nannicini, T. (2018). Regression discontinuity designs based on population thresholds: Pitfalls and solutions. *American Journal of Political Science*, 62(1):210–229.

Fusejima, K., Ishihara, T., and Sawada, M. (2025). A unified test for regression discontinuity designs. *Journal of Econometrics*, 251:106074.

Grembi, V., Nannicini, T., and Troiano, U. (2016). Do fiscal rules matter? *American Economic Journal: Applied Economics*, pages 1–30.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Lalive, R. (2008). How do extended benefits affect unemployment duration? a regression discontinuity approach. *Journal of econometrics*, 142(2):785–806.

Larsen, M. F. and Valant, J. (2023). Fuzzy difference-in-discontinuities when the confounding variation is sharp: Evidence from grade retention policies. *Applied Economics Letters*, 30(15):2009–2013.

Leonardi, M. and Pica, G. (2013). Who pays for it? the heterogeneous wage effects of employment protection legislation. *Economic Journal*, 123(573):1236–1278.

Leventer, D. and Nevo, D. (2024). Correcting invalid regression discontinuity designs with multiple time period data. *arXiv preprint arXiv:2408.05847*.

Manski, C. F. and Pepper, J. V. (2018). How do right-to-carry laws affect crime rates? coping with ambiguity using bounded-variation assumptions. *Review of Economics and Statistics*, 100(2):232–244.

Masten, M. A. and Poirier, A. (2020). Inference on breakdown frontiers. *Quantitative Economics*, 11(1):41–111.

Millán-Quijano, J. (2020). Fuzzy difference in discontinuities. *Applied Economics Letters*, 27(19):1552–1555.

Picchetti, P., Pinto, C. C., and Shinoki, S. T. (2026). Difference-in-discontinuities: estimation, inference and validity tests. *arXiv preprint arXiv:2405.18531*.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of economic studies*, 90(5):2555–2591.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–322.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of econometrics*, 235(2):2218–2244.

Takahashi, M. (2026). A note on identification of difference-in-discontinuities in cross-sectional settings. SSRN Working Paper 4686891, revised April 2026.

Tchuente, G., Galindo-Silva, H., and Somé, N. H. (2026). Fuzzy difference-in-discontinuities: Identification theory and illustration with the Affordable Care Act. *Econometric Reviews*. Published online August 16, 2026.

A Proofs

A.1 Proof of Lemma 1

Proof. We first characterize the observed discontinuity θ_t in each period. For $t \in \mathcal{T}_0$,

$$\begin{aligned}\theta_t &= \lim_{x\to 0^+} \mathbb{E}[Y_{it}(1,0) \mid X_i = x] - \lim_{x\to 0^-} \mathbb{E}[Y_{it}(0,0) \mid X_i = x] \\ &= \mathbb{E}[Y_{it}(1,0) - Y_{it}(0,0) \mid X_i = 0] = \rho_t.\end{aligned}$$

The first equality follows from the assignment rules and the observed-outcome equation; and the second equality follows from assumption 1. Therefore,

$$\beta_t = \theta_t - \theta_0 = \rho_t - \rho_0 = \delta_t, \quad \forall t \in \mathcal{T}_0 \setminus \{0\}.$$

For $t \in \mathcal{T}_1$,

$$\begin{aligned}\theta_t &= \lim_{x\to 0^+} \mathbb{E}[Y_{it}(1,1) \mid X_i = x] - \lim_{x\to 0^-} \mathbb{E}[Y_{it}(0,0) \mid X_i = x] \\ &= \mathbb{E}[Y_{it}(1,1) - Y_{it}(0,0) \mid X_i = 0] \\ &= \mathbb{E}[Y_{it}(1,1) - Y_{it}(1,0) \mid X_i = 0] + \mathbb{E}[Y_{it}(1,0) - Y_{it}(0,0) \mid X_i = 0] \\ &= \tau_t + \rho_t.\end{aligned}$$

The first and second equality follow the same logic as above, and the third equality comes from adding and subtracting the same quantity. Similarly, we get

$$\beta_t = \theta_t - \theta_0 = \tau_t + \rho_t - \rho_0 = \tau_t + \delta_t, \quad \forall t \in \mathcal{T}_1.$$

The result follows from stacking the $\underline{T}$ pre-treatment RD differences and the $\bar{T}$ post-treatment RD differences. □

A.2 Proof of Proposition 1

We first define the limiting covariance matrix. Denote the population conditional covariance between Y_{it} and Y_{iu} on each side of the cutoff as

$$\sigma^2_{tu+} = \lim_{x\to 0^+} \mathrm{Cov}(Y_{it}, Y_{iu} \mid X_i = x), \qquad \sigma^2_{tu-} = \lim_{x\to 0^-} \mathrm{Cov}(Y_{it}, Y_{iu} \mid X_i = x).$$

For $t, u \in \mathcal{T}$, let

$$\begin{aligned}
\Gamma_{+,\ell+1} &= f(0) \int_0^\infty K(v) r_{\ell+1}(v) r_{\ell+1}(v)^\top \, dv, \\
\Gamma_{-,\ell+1} &= f(0) \int_{-\infty}^0 K(v) r_{\ell+1}(v) r_{\ell+1}(v)^\top \, dv, \\
\Psi_{tu+,\ell+1} &= f(0) \sigma^2_{tu+} \sqrt{c_{tu}} \int_0^\infty K(v) K(c_{tu} v) r_{\ell+1}(v) r_{\ell+1}(c_{tu} v)^\top \, dv \\
\Psi_{tu-,\ell+1} &= f(0) \sigma^2_{tu-} \sqrt{c_{tu}} \int_{-\infty}^0 K(v) K(c_{tu} v) r_{\ell+1}(v) r_{\ell+1}(c_{tu} v)^\top \, dv,
\end{aligned}$$

where $c_{tu} = \lim_{n\to\infty} h_{t,n}/h_{u,n}$. The limiting covariance matrix is

$$V_\theta = \left\{ \mathrm{Cov}_{tu+,\ell+1} + \mathrm{Cov}_{tu-,\ell+1} \right\}_{t,u\in\mathcal{T}},$$

where

$$\begin{aligned}
\mathrm{Cov}_{tu+,\ell+1} &= e_0^\top \Gamma^{-1}_{+,\ell+1} \Psi_{tu+,\ell+1} \Gamma^{-1}_{+,\ell+1} e_0, \\
\mathrm{Cov}_{tu-,\ell+1} &= e_0^\top \Gamma^{-1}_{-,\ell+1} \Psi_{tu-,\ell+1} \Gamma^{-1}_{-,\ell+1} e_0.
\end{aligned}$$

Proof. The joint normality result follows from the conditional-mean part of Fusejima et al. (2025, Online Appendix, Theorem A.1) for the outcome vector $(Y_{it} : t \in \mathcal{T})^\top$. The polynomial order is $\ell + 1$, and the centering parameters are the population discontinuities $(\theta_t : t \in \mathcal{T})^\top$. The sampling, moment, smoothness, kernel, and bandwidth conditions are satisfied by imposing Assumptions 2-4 with $S \geq \ell + 3$. The covariance formulas in their Online Appendix A.2, after the change of variable $x = h_{t,n} v$ and using $h_{t,n}/h_{u,n} \to c_{tu}$, give the matrix V_θ defined above. Consequently,

$$\sqrt{n} D_n (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N_T(0, V_\theta).$$

We next establish covariance consistency. Standard local-polynomial moment convergence gives

$$\Gamma_{+,\ell+1}(h_{t,n}) \xrightarrow{p} \Gamma_{+,\ell+1}, \qquad \Gamma_{-,\ell+1}(h_{t,n}) \xrightarrow{p} \Gamma_{-,\ell+1}.$$

Both limiting matrices are positive definite because $f(0) > 0$ and K is positive on $(-1, 1)$. The nearest-neighbor variance consistency argument of Calonico et al. (2014), extended to cross-outcome covariances by Fusejima et al. (2025, Online Appendix A.4), implies under

Assumptions 2-4 that

$$\sqrt{h_{t,n}h_{u,n}}\,\widehat{\Psi}_{tu+,\ell+1}(h_{t,n},h_{u,n}) \xrightarrow{p} \Psi_{tu+,\ell+1},$$
$$\sqrt{h_{t,n}h_{u,n}}\,\widehat{\Psi}_{tu-,\ell+1}(h_{t,n},h_{u,n}) \xrightarrow{p} \Psi_{tu-,\ell+1}.$$

Together with the remaining regularity conditions, the one-sided Lipschitz conditions ensure consistency of the weighted nearest-neighbor covariance products. Combining these limits with convergence of the design matrices and their nonsingular limits, the continuous mapping theorem gives

$$n\sqrt{h_{t,n}h_{u,n}}\,\widehat{\mathrm{Cov}}_{tu+,\ell+1} \xrightarrow{p} \mathrm{Cov}_{tu+,\ell+1},$$
$$n\sqrt{h_{t,n}h_{u,n}}\,\widehat{\mathrm{Cov}}_{tu-,\ell+1} \xrightarrow{p} \mathrm{Cov}_{tu-,\ell+1}\,.$$

Summing the two and collecting the finitely many entries gives $nD_n\widehat{V}_nD_n \xrightarrow{p} V_\theta$. □

A.3 Proof of Corollary 1

Proof. The row of R for period $t \neq 0$ is $e_t^\top - e_0^\top$. Hence the t-th row of $R_n := D_{\beta,n}RD_n^{-1}$ is

$$\sqrt{h_{t,n}}(e_t^\top - e_0^\top)D_n^{-1} = e_t^\top - \sqrt{h_{t,n}/h_{0,n}}\,e_0^\top \to e_t^\top - \sqrt{c_{t0}}\,e_0^\top.$$

The last convergence is given by the bandwidth-ratio condition and continuity of the square root. Since the number of periods is fixed, entrywise convergence also gives matrix convergence, so $R_n \to R_*$.

For each $t \neq 0$, the normalized estimation error is exactly

$$\sqrt{nh_{t,n}}(\widehat{\beta}_t - \beta_t) = \sqrt{nh_{t,n}}(\widehat{\theta}_{t,\ell+1} - \theta_t) - \sqrt{h_{t,n}/h_{0,n}}\,\sqrt{nh_{0,n}}(\widehat{\theta}_{0,\ell+1} - \theta_0).$$

Stacking these equalities in a vector gives

$$\begin{aligned}\sqrt{n}D_{\beta,n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{n}D_{\beta,n}R(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= \sqrt{n}D_{\beta,n}RD_n^{-1}D_n(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= \left(D_{\beta,n}RD_n^{-1}\right)\sqrt{n}D_n(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= R_n\sqrt{n}D_n(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}).\end{aligned}$$

For the sampling covariance,

$$\begin{aligned} nD_{\beta,n}\widehat{\Omega}_n D_{\beta,n} &= nD_{\beta,n}R\widehat{V}_n R^\top D_{\beta,n} \\ &= nD_{\beta,n}RD_n^{-1}\left(D_n\widehat{V}_n D_n\right)D_n^{-1}R^\top D_{\beta,n} \\ &= \left(D_{\beta,n}RD_n^{-1}\right)\left(nD_n\widehat{V}_n D_n\right)\left(D_n^{-1}R^\top D_{\beta,n}\right) \\ &= R_n\left(nD_n\widehat{V}_n D_n\right)R_n^\top, \end{aligned}$$

where the last equality uses the symmetry of the diagonal matrices D_n and $D_{\beta,n}$.

By Proposition 1, $\sqrt{n}D_n(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}_T(0,V_\theta)$ and $nD_n\widehat{V}_nD_n \xrightarrow{p} V_\theta$. Since $R_n \to R_*$, apply Slutsky's theorem,

$$\sqrt{n}D_{\beta,n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}_{T-1}(0, R_*V_\theta R_*^\top).$$

Additionally, by the continuous mapping theorem,

$$nD_{\beta,n}\widehat{\Omega}_n D_{\beta,n} \xrightarrow{p} R_*V_\theta R_*^\top.$$

Denote $V_\beta = R_*V_\theta R_*^\top$, and we conclude the proof. □

Example: three pre-treatment periods and one post-treatment period. Let $\mathcal{T} = \{-2,-1,0,1\}$, with period zero as the baseline. Then

$$\boldsymbol{\theta} = (\theta_{-2},\theta_{-1},\theta_0,\theta_1)^\top, \qquad \boldsymbol{\beta} = \begin{pmatrix}\theta_{-2}-\theta_0\\ \theta_{-1}-\theta_0\\ \theta_1-\theta_0\end{pmatrix}, \qquad R = \begin{pmatrix}1&0&-1&0\\0&1&-1&0\\0&0&-1&1\end{pmatrix}.$$

The normalization matrices are

$$D_n = \operatorname{diag}\left(\sqrt{h_{-2,n}},\sqrt{h_{-1,n}},\sqrt{h_{0,n}},\sqrt{h_{1,n}}\right), \qquad D_{\beta,n} = \operatorname{diag}\left(\sqrt{h_{-2,n}},\sqrt{h_{-1,n}},\sqrt{h_{1,n}}\right).$$

Consequently,

$$R_n = D_{\beta,n}RD_n^{-1} = \begin{pmatrix}1&0&-\sqrt{h_{-2,n}/h_{0,n}}&0\\0&1&-\sqrt{h_{-1,n}/h_{0,n}}&0\\0&0&-\sqrt{h_{1,n}/h_{0,n}}&1\end{pmatrix} \longrightarrow R_* = \begin{pmatrix}1&0&-\sqrt{c_{-2,0}}&0\\0&1&-\sqrt{c_{-1,0}}&0\\0&0&-\sqrt{c_{1,0}}&1\end{pmatrix}.$$

To express the limiting distribution coordinate by coordinate, define

$$\zeta_{t,n} = \sqrt{nh_{t,n}}\{\widehat{\theta}_{t,\ell+1}(h_{t,n}) - \theta_t\}.$$

Proposition 1 gives

$$\boldsymbol{\zeta}_n = (\zeta_{-2,n}, \zeta_{-1,n}, \zeta_{0,n}, \zeta_{1,n})^\top \xrightarrow{d} \boldsymbol{\zeta} = (\zeta_{-2}, \zeta_{-1}, \zeta_0, \zeta_1)^\top \sim \mathcal{N}_4(0, V_\theta).$$

Therefore,

$$\begin{aligned}
\sqrt{n}D_{\beta,n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \begin{pmatrix} \sqrt{nh_{-2,n}}(\widehat{\beta}_{-2} - \beta_{-2}) \\ \sqrt{nh_{-1,n}}(\widehat{\beta}_{-1} - \beta_{-1}) \\ \sqrt{nh_{1,n}}(\widehat{\beta}_{1} - \beta_{1}) \end{pmatrix} \\
&= \begin{pmatrix} \zeta_{-2,n} - \sqrt{h_{-2,n}/h_{0,n}}\,\zeta_{0,n} \\ \zeta_{-1,n} - \sqrt{h_{-1,n}/h_{0,n}}\,\zeta_{0,n} \\ \zeta_{1,n} - \sqrt{h_{1,n}/h_{0,n}}\,\zeta_{0,n} \end{pmatrix} \\
&\xrightarrow{d} \begin{pmatrix} \zeta_{-2} - \sqrt{c_{-2,0}}\,\zeta_0 \\ \zeta_{-1} - \sqrt{c_{-1,0}}\,\zeta_0 \\ \zeta_{1} - \sqrt{c_{1,0}}\,\zeta_0 \end{pmatrix} = R_*\boldsymbol{\zeta} \sim \mathcal{N}_3(0, V_\beta),
\end{aligned}$$

where $V_\beta = R_* V_\theta R_*^\top$. Each difference contains its own period's estimation error and the common baseline error, and the latter is rescaled by the corresponding bandwidth ratio. Equal bandwidths give $R_* = R$.

A.4 Proof of Proposition 2

Proof. We verify the assumptions of Rambachan and Roth (2023, Proposition 3.1), which applies the size control result of Andrews et al. (2023, Proposition 2) for each fixed data-generating process, allowing for the convergence rate of the local-polynomial estimators.

1. Asymptotic normality. Let

$$a_n = \sqrt{nh_{0,n}}, \qquad G_n = \sqrt{h_{0,n}}D_{\beta,n}^{-1}, \qquad G = \operatorname{diag}\{c_{t0}^{-1/2} : t \neq 0\}.$$

Our Assumption 4 implies $a_n \to \infty$ and $G_n \to G$, with G nonsingular. By Corollary 1 and Slutsky's theorem,

$$a_n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = G_n\sqrt{n}D_{\beta,n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}_{T-1}(0, V), \qquad V = GV_\beta G^\top. \tag{23}$$

Weak convergence implies convergence in the bounded-Lipschitz metric, given their normal-approximation condition at rate a_n. Although Rambachan and Roth (2023, Proposition 3.1) formulate their statistical assumptions using a $\sqrt{n}$ normalization, the result extends to any deterministic rate $a_n \to \infty$ satisfying the corresponding normal approximation and covariance consistency conditions. To see this, along any subsequence, select a further subsequence for which $m_j = \lfloor a_{n_j}^2 \rfloor$ is strictly increasing. Reindex the original estimator pairs at m_j and complete the sequence at the remaining indices with Gaussian estimates having mean $\boldsymbol{\beta}$ and covariance V/m. The resulting sequence satisfies their assumptions with normalization $\sqrt{m}$, while preserving the original test inputs at m_j. Their size control result therefore applies to the original sequence, including along sequences of candidate values in the identified set.

2. Eigenvalue bounds. The matrix R_* has full row rank because its nonbaseline columns form the identity matrix. Since V_θ is a nonsingular covariance matrix, it is positive definite. Thus, for every nonzero $x \in \mathbb{R}^{T-1}$,

$$x^\top V x = (R_*^\top G^\top x)^\top V_\theta (R_*^\top G^\top x) > 0.$$

Here $R_*^\top G^\top x \neq 0$ by the rank of R_* and nonsingularity of G. Hence V is positive definite. As P is fixed, its eigenvalues admit fixed positive lower and finite upper bounds, as required.

3. Covariance consistency. Corollary 1 and the continuous mapping theorem give

$$a_n^2 \widehat{\Omega}_n = G_n \big(n D_{\beta,n} \widehat{\Omega}_n D_{\beta,n} \big) G_n^\top \xrightarrow{p} G V_\beta G^\top = V. \tag{24}$$

Thus, $a_n^2 \widehat{\Omega}_n$ consistently estimates the normalized covariance, while $\widehat{\Omega}_n$ remains the sampling covariance used in the tests.

4. Restriction geometry. Our Assumption 5 is precisely their Assumption 5(A). The restriction $\Delta = \{\delta : A\delta \leq d\}$ is fixed and nonempty, and A has nonzero rows, as maintained in Section 6.

These conditions hold uniformly over all admissible decompositions $\boldsymbol{\beta} = \boldsymbol{\delta} + L_{\text{post}} \boldsymbol{\tau}_{\text{post}}$ with $\delta \in \Delta$ under the fixed observable distribution P. This is because all admissible decompositions induce the same distribution of the estimators. Thus, we have verified all the assumptions to apply Proposition 3.1 in Rambachan and Roth (2023) for size control of the tests, and therefore conclude the proof.

□

B Simulation Design

B.1 Data-generating processes

For each of 2000 replications per design, we draw $n = 1000$ independent units and observe each unit at $t = \{-2, -1, 0, 1\}$. Let $X_i = 2B_i - 1$, where $B_i \sim \text{Beta}(2, 4)$, and set the cutoff to zero. The running variable is fixed across periods. Outcomes satisfy

$$Y_{it} = 0.5 + a_t + s_t g(X_i) + \mathbf{1}\{X_i \geq 0\}\big(\rho_t + 0.20\,\mathbf{1}\{t = 1\}\big) + \varepsilon_{it},$$

where,$(a_{-2}, a_{-1}, a_0, a_1) = (-0.10, -0.05, 0, 0.05)$, and $(s_{-2}, s_{-1}, s_0, s_1) = (0.80, 0.90, 1, 1.15)$. The continuous function g uses the same polynomial coefficients in Picchetti et al. (2026, Appendix D.2.2):

$$\begin{aligned}
g(x) &= \begin{cases} g_-(x), & x < 0, \\ g_+(x), & x \geq 0, \end{cases} \\
g_-(x) &= 1.27x - 0.5(7.18)x^2 + 0.7(20.21)x^3 \\
&\quad + 1.1(21.54)x^4 + 1.5(7.33)x^5, \\
g_+(x) &= 0.84x - 0.1(3)x^2 - 0.3(7.99)x^3 - 0.1(9.01)x^4 + 3.56x^5.
\end{aligned}$$

Every discontinuity comes from the explicitly specified confounding and treatment effects. The level shifts a_t and scale factors s_t allow conditional means to change over time without changing their cutoff jumps.

Independently of X_i, draw the four-dimensional error vector from $\mathcal{N}(0, \Sigma_\varepsilon)$, where

$$[\Sigma_\varepsilon]_{ts} = 0.1295^2\, 0.5^{|t-s|}, \qquad t, s \in \{-2, -1, 0, 1\}.$$

Thus, errors are independent across units, have the same marginal variance in every period, and have AR(1) correlation within units. The design uses the running-variable distribution and marginal error standard deviation of Picchetti et al. (2026), with four periods and serially correlated errors.

The two designs differ only in the confounding effects for each period. The first has $\rho_t = 0.30$ in every period. The second has the vector of ρ_t as $(0.10, 0.18, 0.28, 0.43)^\top$. We observe $\theta_t = \rho_t$ before treatment and $\theta_1 = \rho_1 + 0.20$ afterwards. Applying the differencing with respect to the last pre-treatment period, $\boldsymbol{\beta} = (\theta_{-2} - \theta_0, \theta_{-1} - \theta_0, \theta_1 - \theta_0)^\top$ is $(0, 0, 0.20)^\top$ in the first design and $(-0.18, -0.10, 0.35)^\top$ in the second. The target is $\vartheta = \tau_1 = 0.20$, with scalar weight $l = 1$.

B.2 Sensitivity restrictions and population benchmarks

We evaluate RM on a grid from $\bar{M} = 0$ to 3 in increments of 0.5, and SD on a grid from $M = 0$ to 0.15 in increments of 0.025. For these four periods, RM bounds $|\rho_1 - \rho_0|$ by $\bar{M}$ times $\max\{|\rho_{-1} - \rho_{-2}|, |\rho_0 - \rho_{-1}|\}$. SD bounds both second differences, $|\rho_0 - 2\rho_{-1} + \rho_{-2}|$ and $|\rho_1 - 2\rho_0 + \rho_{-1}|$, by M. The constant path satisfies both restrictions at every nonnegative bound. For the increasing path, the first differences are $(0.08, 0.10, 0.15)$ and the second differences are $(0.02, 0.05)$. The generating path therefore satisfies RM exactly when $\bar{M} \geq 1.5$, and SD exactly when $M \geq 0.05$.

The population identified sets follow directly by substituting $\tau_1 = \beta_1 - (\rho_1 - \rho_0)$ into the restrictions. Under constant confounding, they are

$$\mathcal{S}(\boldsymbol{\beta}, \Delta^{\mathrm{RM}}(\bar{M})) = \{0.20\}, \qquad \mathcal{S}(\boldsymbol{\beta}, \Delta^{\mathrm{SD}}(M)) = [0.20 - M, 0.20 + M].$$

Under increasing confounding they are

$$\mathcal{S}(\boldsymbol{\beta}, \Delta^{\mathrm{RM}}(\bar{M})) = [0.35 - 0.10\bar{M},\ 0.35 + 0.10\bar{M}],$$
$$\mathcal{S}(\boldsymbol{\beta}, \Delta^{\mathrm{SD}}(M)) = \begin{cases} \varnothing, & 0 \leq M < 0.02, \\ [0.25 - M,\ 0.25 + M], & M \geq 0.02. \end{cases}$$

For SD, the empty case reflects incompatibility with the observed pre-treatment curvature. When this curvature is allowed, SD bounds the post-treatment change around the last pre-treatment change, 0.10, which gives the center $0.35 - 0.10 = 0.25$. These calculations distinguish compatibility with the pre-treatment path from inclusion of the generating effect, which requires $M \geq 0.05$. Picchetti et al. (2026)'s population estimand is $\theta_1 - \theta_0$, equal to 0.20 and 0.35 in the two designs.

B.3 Estimation and confidence set construction

The raw RD plots use local linear estimation with quadratic bias correction, the same estimation and pilot bandwidth, a triangular kernel, and three-nearest-neighbor variance estimation. The Picchetti et al. (2026) confidence interval applies `rdrobust` directly to the differenced outcome, $Y_{i1} - Y_{i0}$, with the same polynomial orders, kernel, and variance method. Its estimation and bias bandwidths are selected separately using the package defaults for `mserd`.

For constructing confidence sets under restrictions $\Delta^{\mathrm{RM}}(M)$ and $\Delta^{\mathrm{SD}}(M)$, we obtain the period-specific higher-order estimates and their full sampling covariance from the implemen-

tation of Fusejima et al. (2025) in their package `rdtest`.[1] This implementation selects a common bandwidth across the two sides within each period, allows bandwidths to differ across periods, and uses local quadratic intercepts at the bandwidths selected for local linear estimation. The covariance uses common nearest neighbors across outcomes to preserve within-unit dependence. We transform the estimated $\widehat{\boldsymbol{\theta}}$ and its covariance as discussed in Section 5 to get $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\Omega}_n$. The covariance supplied to inference is the sampling covariance, without an additional division by n.

We pass $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\Omega}_n$ to the publicly available RM and SD sensitivity functions in `HonestDiD`, with two pre-treatment coefficients, one post-treatment coefficient, $l = 1$, $\alpha = 0.05$, and `method="C-LF"`. We demonstrate our simulation results using the suggested hybrid confidence sets by Rambachan and Roth (2023). All remaining tuning and search-grid options use package defaults. The implementation makes one call per sensitivity value and obtains the returned bounds.

[1] The source code of the `rdtest` package can be found at https://github.com/SMasa11/rdtest.