DIFFERENCE-IN-DISCONTINUITIES: ESTIMATION, INFERENCE AND VALIDITY TESTS

Pedro Picchetti[*]
pedrop3@al.insper.edu.br

Cristine C. X. Pinto [*]
cristinecxp@insper.edu.br

Stéphanie T. Shinoki[*]
stephaniets@al.insper.edu.br

ABSTRACT

This paper investigates the econometric theory behind the newly developed difference-in-discontinuities design (DiDC). Despite its increasing use in applied research, there are currently limited studies of its properties. The method combines elements of regression discontinuity (RDD) and difference-in-differences (DiD) designs, allowing researchers to eliminate the effects of potential confounders at the discontinuity. We formalize the difference-in-discontinuity theory by stating the identification assumptions and proposing a nonparametric estimator, deriving its asymptotic properties and examining the scenarios in which the DiDC has desirable bias properties when compared to the standard RDD. We also provide comprehensive tests for one of the identification assumption of the DiDC. Monte Carlo simulation studies show that the estimators have good performance in finite samples. Finally, we revisit Grembi et al. (2016), that studies the effects of relaxing fiscal rules on public finance outcomes in Italian municipalities. The results show that the proposed estimator exhibits substantially smaller confidence intervals for the estimated effects.

Keywords: Difference-in-discontinuities, regression discontinuity design, difference-in-difference

1 Introduction

The difference-in-discontinuities design (DiDC) aims to solve the limitations of both regression discontinuity designs (RDD) and difference-in-difference designs (DiD) by combining both the temporal and discontinuity-based sources of variation from the data-generating processes (DGP). Grembi et al. (2016) and Eggers et al. (2018) proposed this quasi-experimental approach, examining the difference between pre- and post-treatment periods around a threshold for both treated and not treated groups. Although its use in applied microeconomics works is increasing (Azuaga and Sampaio, 2017; Chicoine, 2017; García, 2020; Albright, 2023), research on the econometric theory of the DiDC, including identification assumptions, estimation, and asymptotic properties, remains limited.

The method offers more flexibility over the standard RDD in some settings and can often be used in contexts where neither the RDD nor the DiD is applicable. In particular, the DiDC can handle cases where control and treatment groups differ significantly so as not to satisfy the parallel trends assumption of the DiD, or when multiple time-invariant confounders exist at the threshold in an RDD setting. Additionally, by incorporating more information into the estimation, the DiDC eliminates the bias in the RDD estimates providing more accurate and reliable estimates of treatment effects under certain assumptions, including ones on the DGP's functional form.

In this paper, we develop the econometric theory for the design with a discontinuity-based approach to the differences. While Galindo-Silva et al. (2021) develop an identification theory for the *fuzzy* difference-in-discontinuity design based on the difference of RDD estimations, this work will focus on the complementary approach of studying the discontinuity of the differences.

Drawing from standard assumptions in cross-sectional RDD and assumptions specific to the DiDC framework, we establish conditions for reliable identification. These new assumptions, specific to the DiDC design, restrict how the confounding effects behave around the threshold and through time.

[*]Department of Economics, Insper Institute of Education and Research, Rua Quatá, 300, São Paulo, SP, Brazil 04546-042.

We show that the parameter of interest can be recovered using local polynomial estimation of the differences in the outcomes employing the methodology proposed by Calonico et al. (2014). We derive the asymptotic properties of the estimator, highlighting scenarios where the asymptotic bias of the DiDC can be smaller than that of the RDD. Additionally, we introduce straightforward tests to evaluate the testable implications of necessary assumptions for identification. We explore the conditions under which DiDC mitigates bias more effectively than the RDD offering practical guidance for unbiased treatment effect estimation.

Through Monte Carlo simulations, we evaluate the finite-sample properties of the estimator comparing its performance to that of the local linear RDD estimator proposed by Calonico et al. (2014) and the nonparametric DiD regression estimator proposed by Sant'Anna and Zhao (2020). By comparing DiDC's performance with DiD and RDD methods, we identify the scenarios where DiDC outperforms RDDs and we determine the appropriate application contexts for each method. Finally, we apply our estimator to the study by Grembi et al. (2016) on the impact of fiscal rules on municipal deficits in Italy.

This work is closely connected to Calonico et al. (2014), whose contributions have been instrumental to robust nonparametric RDD estimation and whose estimation methods are used for our DiDC approach. Frölich and Sperlich (2019) has a section exploring the possibilities of an intersection between RDD and DiD however, the absence of a formal derivation of econometric properties leaves a gap in understanding, with only preliminary discussions on identification and estimation procedures. This work also aligns with the broader literature on the intersection of RDD and panel data. Pettersson-Lidbom (2012) explores settings where RDD is combined with fixed effects to address small sample issues and violations of the continuous support assumption. Lemieux and Milligan (2008), use a first-difference RD approach to eliminate individual-specific fixed effects by capitalizing on the longitudinal nature of the Finnish Census data. Lastly, Cellini et al. (2010) introduce "dynamic RD" models, accommodating scenarios with multiple treatment opportunities and looking into the dynamics of treatment effects.

The advantages of the DiDC have been explored across several applied microeconomic studies. For instance, Azuaga and Sampaio (2017) uses the DiDC to analyze the impact of introducing legislation on domestic violence in Brazil. Butts (2021) explore the use of Diff-in-Disc in geographic settings, Chicoine (2017) investigates the expiration of the Assault Weapon Ban (AWB) by comparing municipalities in which the incumbent mayoral party wins a close election with those where the incumbent is defeated and Albright (2023) isolates the causal effects of algorithmic recommendations on decision-makers using DiDC. These real-life examples demonstrate how effective this approach can be in various research settings, highlighting its value as a practical analysis method.

The remainder of this article is organized as follows. Section 2 presents the DiDC as the discontinuity of differences in a potential outcomes model and shows the main identification results, along with the necessary assumptions. Estimation procedures for the treatment effect in the *sharp* setting are presented in Section 3, along with the derivation of large-sample properties, optimal bandwidths and robust confidence intervals. In Section 4 we present tests for 2 of the identifying assumptions. Monte Carlo simulations are conducted in Section 5 to examine the properties of the estimators. Section 6 provides an empirical illustration and Section 7 concludes. In the Appendix we provide detailed notation, proofs and other methodological results.

2 Difference-in-Discontinuities as the discontinuity of differences

2.1 Setup

We define the potential outcomes as a function of both treatment assignment and time. There are two time periods, $t = 0$ and $t = 1$, with treatment determined by a threshold z_0 on the running variable Z_i at $t = 0$. Treatment is introduced sometime between $t = 0$ and $t = 1$ so that no units are treated by the treatment of interest in $t = 0$. If there's a pre-existing confounder due to the discontinuity at z_0 (a confounding and time-invariant discontinuity in the outcomes at $t = 0$), we refer to it as the "confounding treatment". We indicate whether units were "treated" in period t with $D_{i,t} = 1$ and units that were not "treated" with $D_{i,t} = 0$, where $D_{i,0} = 1$ implies "treatment" by the confounding policy and $D_{i,1} = 1$ implies treatment by the actual treatment of interest. We assume that the assignment rule and the running variables remain constant throughout both periods.

We define four potential outcomes ($Y_{i,t}(D_1, D_0)$) for each unit at each t, which depend on the treatment of interest and the confounding treatment:

$$Y_{i,t}(0,0) \quad Y_{i,t}(0,1) \quad Y_{i,t}(1,0) \quad Y_{i,t}(1,1)$$

We are interested in the local causal effect of the treatment of interest, $D_1 = 1$, at $Z = z_0$. To estimate this we compare the difference in outcomes between pre- and post-treatment periods for units near the threshold, both for treated and untreated groups. This comparison can be approached through two methods: the differences of RDDs and the RDD of the differences. While prior studies have predominantly employed the former, this research opts for the latter approach. This choice is because the RDD of the differences approach offers a closer analogy to the well-established and robust RDD literature. However, it is essential to highlight that both methods lead to the same estimand, ultimately identifying the same underlying treatment effects, as demonstrated in Claim 1 shown in the next section.

The observed outcome under the perspective of the difference of RDDs is:

$$Y_{i,1} = \left[Y_{i,1}(1,1)\,D_0 + Y_{i,1}(0,1)\,(1-D_0)\right]D_1 + \left[Y_{i,1}(1,0)\,D_0 + Y_{i,1}(0,0)\,(1-D_0)\right](1-D_1) \tag{1}$$
$$Y_{i,0} = Y_{i,0}(1,0)\,D_0 + Y_{i,0}(0,0)\,(1-D_0) \tag{2}$$

Where $Y_{i,0}$ takes this form because there is no treatment D_1 at time $t = 0$.

Meanwhile, the observed outcome under the approach of the RDD of the differences between outcomes at times $t = 0$ and $t = 1$ is:

$$\begin{aligned} \Delta Y_i &= \Delta Y_i(0) + \left[\Delta Y_i(1) - \Delta Y_i(0)\right]D_1 \\ \Delta Y_i\,(D_1) &= Y_{i,1}\,(D_1) - Y_{i,0} \\ Y_{i,1}\,(D_1) &= Y_{i,1}\,(1,D_1) + \left[Y_{i,1}\,(1,D_1) - Y_{i,1}\,(0,D_1)\right]D_0 \\ Y_{i,0} &= Y_{i,0}\,(0,0) + \left[Y_{i,0}\,(1,0) - Y_{i,0}\,(0,0)\right]D_0 \end{aligned} \tag{3}$$

2.2 Identification

Identification in difference-in-discontinuities design relies on the standard assumptions from the cross-sectional RDD given by Hahn et al. (2001) (Assumptions 1 and 2, adapted to include two time-periods), and assumptions concerning the behavior of confounding effects over time (Assumptions 3 and 4). We state the assumptions here.

Assumption 1 (Continuity)**.** *All potential outcomes are continuous in $Z = z_0$, for any $D_0, D_1 \in \{0,1\}$ and $t \in \{0,1\}$:*

$$\lim_{\varepsilon\to 0} E[Y_{it}(D_0,D_1)|Z_i = z_0+\varepsilon] = \lim_{\varepsilon\to 0} E[Y_{it}(D_0,D_1)|Z_i = z_0-\varepsilon]$$

Assumption 2 (Discontinuity of treatment probability)**.** *The limits $D_t^+ = \lim_{\epsilon\to 0} E(D_{it}|Z_{it} = z_0+\epsilon)$ and $D_t^- = \lim_{\epsilon\to 0} E(D_{it}|Z_{it} = z_0-\epsilon)$ exist and $D_t^+ \neq D_t^-$.*

Assumption 2 is a relevance condition, which needs to hold for identification in both *sharp* and *fuzzy* settings. In the *sharp* settings, however, it simplifies to $D_1^+ = 1$ and $D_1^- = 0$, while $D_0^+ = D_0^- = 0$.

Claim 1. Under Assumptions 1 and 2, the difference of RDDs, and the RDD of the differences are the same

$$\tau^{DiDC} = \tau_1^{RDD} - \tau_0^{RDD} = \lim_{\varepsilon\to 0} E[\Delta Y_t|Z_{it} = z_0+\epsilon] - \lim_{\varepsilon\to 0} E[\Delta Y_t|Z_{it} = z_0-\epsilon]$$

Proof. See Appendix C. ☐

To establish the identification of the treatment effect, $E[Y_i(1) - Y_i(0)|Z_i = z_0]$ we invoke Assumptions 3 and 4, which are specific to the DiDC design and concern the behavior of the confounding effects around the threshold.

Assumption 3 (Time-invariance of confounding effects)**.** *The treatment of interest is the only time-variant effect at the threshold z_0:*

$$E\left[Y_{i,0}(1,0) - Y_{i,0}(0,0)|Z_i = z_0\right] = E\left[Y_{i,1}(1,0) - Y_{i,1}(0,0)|Z_i = z_0\right]$$

Assumption 3 assumption asserts that the effect of confounding policy D_0 when there is no treatment remains constant over time. That is, any differences between the treatment and control groups at the threshold in $t = 1$, not caused by the treatment, should have existed in $t = 0$ before the treatment was introduced. The importance of this assumption cannot be understated. Its validity directly affects the credibility of all subsequent estimates and without it, the estimates from DiDC analyses are unreliable and biased. In Section 4.1, we introduce a comprehensive test to assess its validity.

Assumption 4 (Treatment effect is independent from the confounding policy)**.**

$$E\left[Y_{i,t}(1,1) - Y_{i,t}(1,0)|Z_i = z_0\right] = E\left[Y_{i,t}(0,1) - Y_{i,t}(0,0)|Z_i = z_0\right] = E\left[Y_{i,t}(1) - Y_{i,t}(0)|Z_i = z_0\right]$$

This final assumption needed for identification states that the treatment effect should not be affected by the confounding policy. This condition implies that the total effect captured by an RDD at time $t = 1$ would be the sum of two effects: the effect of the confounding treatment and the effect of the treatment of interest. That is, there is no complementarity or interaction between the treatment of interest and the confounding treatment.

Under Assumptions 1 - 4, we derive the *Average Treatment Effect* of treatment D_1:

$$\begin{aligned}\tau^{DiDC} &= \lim_{\varepsilon\to 0} E[\Delta Y_t|Z_{it} = z_0 + \epsilon] - \lim_{\varepsilon\to 0} E[\Delta Y_t|Z_{it} = z_0 - \epsilon] \\ &= E\left[Y_i\left(1\right) - Y_i\left(0\right)|Z_i = z_0\right]\left(D_1^+ - D_1^-\right)\end{aligned}$$

Therefore the estimand for the DiDC is:

$$\hat{\tau}^{DiDC} = \frac{\Delta Y^+ - \Delta Y^-}{\left(D_1^+ - D_1^-\right)} \tag{4}$$

2.2.1 Relaxing the Independence Assumption

Assumption 4 can be overly restrictive and sometimes there could be scenarios where the confounding policy and the treatment of interest might interact. Rather than assuming additive effects at time $t = 1$, we can relax this assumption with one that allows the inclusion of multiplicative effects:

Assumption 4' (Multiplicative Effects)**.**

$$\begin{aligned}&E\left[Y_{i,t}(1,1) - Y_{i,t}(1,0) - Y_{i,t}(0,1) + Y_{i,t}(0,0)|Z_i = z_0\right]\left(D_0^+D_1^+ - D_0^-D_1^-\right) \\ &\quad = E\left[Y_{i,t}(0,1) - Y_{i,t}(0,0)|Z_i = z_0\right]\left(D_1^+ - D_1^-\right)E\left[Y_{i,t}(1,0) - Y_{i,t}(0,0)|Z_i = z_0\right]\left(D_0^+ - D_0^-\right)\end{aligned}$$

This assumption allows for the possibility that the combined effect of the confounding policy and the treatment of interest is not simply the sum of their individual effects but it also includes an additional component that reflects how both treatments might interact. This additional component is determined by multiplying the effects together. By incorporating this concept, we can derive a new estimand that uses both the RDD at time 0 and the DiDC to identify the causal effect of the treatment of interest:

$$\frac{\tau^{DiDC}}{1 + \tau_0^{RDD}} = E\left[Y_{i,t}(0,1) - Y_{i,t}(0,0)|Z_i = z_0\right]\left(D_1^+ - D_1^-\right)$$

This estimand can capture more complex relationships between the treatment and the confounding factors. Our future work will focus on the properties and robustness of this estimator.

3 Estimation and Inference for the *Sharp* DiDC

Following standard practice in the regression discontinuity literature, we adopt local polynomial estimation to recover the parameter of interest. This nonparametric method involves fitting a polynomial function to the data near the threshold and using the estimated function to calculate the differences in outcomes between the treatment and control groups at the threshold.

To implement the estimation at z_0, we lean on the methodologies proposed by Calonico et al. (2014) for local polynomial estimation of the RDD estimator. In our case, we estimate the local polynomial regression on the differences in the outcomes with respect to time (ΔY_i).

3.1 Notation

Given a known threshold value z_0, which can be set to $z_0 = 0$ without loss of generality, the observed value Z_i determines treatment assignment, with individuals assigned to the treatment group if $Z_i \geq 0$ and to the control group if $Z_i < 0$ in the *sharp* setting.

The random variables $Y_{i,t}(1)$ and $Y_{i,t}(0)$ denote the potential outcomes of unit i with and without treatment, respectively, at period t. The parameter of interest is $E\left[Y_i\left(1\right) - Y_i\left(0\right)|Z_i = z_0\right]$, the average treatment effect of treatment D_1 at the threshold. Under a mild continuity condition, Hahn et al. (2001) showed that $\tau^{RDD} = E[Y_i(1) - Y_i(0)|Z = z_0]$ is nonparametrically identifiable as the difference of two conditional expectations evaluated at the (induced) boundary point $z_0 = 0$. Similarly, the *sharp* DiDC parameter can be identified as the difference of the difference (in time) of two conditional expectations evaluated at $z_0 = 0$ at each side of z_0:

$$\begin{aligned}\tau^{DIDC} &= \Delta\mu_+ - \Delta\mu_-, \\ \Delta\mu_+ &= \lim_{z\to 0^+}\Delta\mu(z), \quad \Delta\mu_- = \lim_{z\to 0^-}\Delta\mu(z) \\ \Delta\mu(z) &= E\left[\Delta Y_i|Z_i = z\right]\end{aligned}$$

We employ Assumptions A.1 and A.2 in Appendix A for the non-parametric local polynomial regression estimation. These assumptions impose restrictions on the kernel function as well as impose the existence of moments, enforce continuity of the running variable in the region, impose smoothness conditions on the regression functions, and restrict the conditional variance of the observed outcome.

3.2 Local Polynomial Estimator

For a given $\nu \leq p \in \mathbf{N}$, the general estimand of interest is $\tau_{\nu}^{DIDC} = \Delta\mu_{+} - \Delta\mu_{-}$ with $\Delta\mu_{+}^{(\nu)} = \nu! e_{\nu}' \delta_{+,p}$, $\Delta\mu_{-}^{(\nu)} = \nu! e_{\nu}' \delta_{-,p}$ being the νth-order derivatives of the pth-order local polynomial of the difference. The pth-order local polynomial estimators of the νth-order derivatives $\Delta\mu_{+,p}^{(\nu)}$ and $\Delta\mu_{-,p}^{(\nu)}$ are:

$$
\begin{aligned}
\Delta\hat{\mu}_{+,p}^{(\nu)}(h_n) &= \nu! e_{\nu}' \hat{\delta}_{+,p}(h_n) \\
\Delta\hat{\mu}_{-,p}^{(\nu)}(h_n) &= \nu! e_{\nu}' \hat{\delta}_{-,p}(h_n) \\
\hat{\delta}_{\Delta Y+,p}(h_n) &= arg \min_{\delta \in \mathbf{R}^{p+1}} \sum_{i=1}^{n} \mathbb{1}(Z_i \geq 0)(\Delta Y_i - r_p(Z_i)'\delta)^2 K_{h_n}(Z_i) \\
\hat{\delta}_{\Delta Y-,p}(h_n) &= arg \min_{\delta \in \mathbf{R}^{p+1}} \sum_{i=1}^{n} \mathbb{1}(Z_i < 0)(\Delta Y_i - r_p(Z_i)'\delta)^2 K_{h_n}(Z_i)
\end{aligned}
$$

where e_{ν} is a conformable $(\nu + 1)$ unit vector, $K_h(u) = K(u/h)/h$, h_n is a positive bandwidth sequence, $r_p(x) = [1 \quad x \quad \ldots \quad x^p]'$ and $\Delta Y = [\Delta Y_1 \quad \Delta Y_2 \quad \ldots \quad \Delta Y_n]'$. Therefore, for a positive bandwidth h_n, the nonparametric estimator of $\tau_{\nu,p}$ is

$$
\hat{\tau}_{\nu,p}^{DiDC}(h_n) = \Delta\hat{\mu}_{+,p}^{(\nu)}(h_n) - \Delta\hat{\mu}_{-,p}^{(\nu)}(h_n) \tag{5}
$$

3.2.1 Bandwidth choice

In order to perform the local polynomial estimation, it is necessary to choose an appropriate bandwidth h_n. This parameter determines the range of observations used for the estimation and impacts the trade-off between bias and variance in the estimated treatment effect (τ^{DiDC}). When performing point estimation, the standard approach is to find the bandwidth that minimizes the asymptotic Mean Squared Error (MSE) of the estimator. We define the MSE as:

$$
MSE_{\nu,p,s}(h_n) = E\left[\left(\hat{\tau}_{\nu,p,s}(h_n) - \tau_{\nu,p} \right)^2 | \chi_n \right]
$$

Lemma 1. Under Assumptions 1 and 2 with $S \geq p+1$, $\nu \leq p$, $h_n \to 0$ and $nh_n \to \infty$, the asymptotic MSE-optimal bandwidth is given by:

$$
h_n^{MSE}{}_{\nu,p} = \left(\frac{(1+2\nu)\,\mathbf{V}_{\nu,p}}{n2\,(1+p-\nu)\,\mathbf{B}_{\nu,p,p+1,s}^2} \right)^{\frac{1}{2p+3}}
$$

where $\mathbb{V}_{\nu,p} = \nu! \frac{\sigma_+^2 - \sigma_-^2}{f} e_{\nu}' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1}$ and $\mathbb{B}_{\nu,p,p+1,s} = \frac{\Delta\mu_+^{(p+1)} - (-1)^{\nu+p+s} \Delta\mu_-^{(p+1)}}{(p+1)!} \nu! e_{\nu}' \Gamma_p^{-1} \varphi_{p,p+1}$, provided that $\mathbf{B}_{\nu,p,p+1,s} \neq 0$.

Proof. in Appendix C. □

3.3 Inference

We now discuss the asymptotic properties of the estimator. Using Lemma B.4 found in the Appendix B, it is possible to recover the leading asymptotic bias, expressed as:

$$
E\left[\hat{\tau}_{\nu,p}^{DiDC}(h_n)|\chi_n\right] - \tau_{\nu} = h_n^{p+1-\nu}\mathbf{B}_{\nu,p,p+1}(h_n) + h_n^{p+2-\nu}\mathbf{B}_{\nu,p,p+2}(h_n) \tag{6}
$$

$$
\mathbf{B}_{\nu,p,r}(h_n) = \frac{\Delta\mu_+^{(r)} \mathcal{B}_{+,\nu,p,r}(h_n) - \Delta\mu_-^{(r)} \mathcal{B}_{-,\nu,p,r}(h_n)}{r!}
$$

where $\mathcal{B}_{+,\nu,p,r} = \nu! e_\nu' \Gamma_{+,p}^{-1}(h_n)\vartheta_{+,p,r}(h_n)$ and $\mathcal{B}_{-,\nu,p,r} = \nu! e_\nu' \Gamma_{-,p}^{-1}(h_n)\vartheta_{-,p,r}(h_n)$ are asymptotically bounded. Further notation is available in Appendix A and a detailed proof for the statement above can be found in Appendix B under Lemma B.4.

This is where we believe is one of the main contributions of this research: the asymptotic bias of the DiDC can be zero if we include an assumption similar to that of parallel trends, or if the shapes of the data-generating processes for both groups are time-invariant. In the first case, imposition of "parallel trends" restricts the functional form in such a way that $\Delta\mu_+^{(r)}\mathcal{B}_{+,\nu,p,r}(h_n) = \Delta\mu_-^{(r)}\mathcal{B}_{-,\nu,p,r}(h_n)$. It's important to note that the symmetry of the kernel function, as imposed in Assumption A.2, plays a significant role in this result. In the case of time-invariant data-generating processes, both $\Delta\mu_+^{(\nu)}$ and $\Delta\mu_-^{(\nu)}$ equate to zero.

We present Claim 2, demonstrating the bias of difference-in-discontinuities in relation to RDDs:

Claim 2. The bias of $\hat{\tau}_{\nu,p}^{DiDC}$ can be decomposed as

$$B\left[\hat{\tau}_{\nu,p}^{DiDC}(h_n)\right] = B\left[\hat{\tau}_{1,\nu,p}^{RDD}(h_n)\right] - B\left[\hat{\tau}_{0,\nu,p}^{RDD}(h_n)\right] \tag{7}$$

where $B\left[\hat{\tau}_{1,\nu,p}^{RDD}(h_n)\right]$ is the bias of the RDD estimated at time $t = 1$, after the intervention happened, and $B\left[\hat{\tau}_{0,\nu,p}^{RDD}(h_n)\right]$ is the bias of the RDD estimated at time $t = 0$, before the intervention happened.

Proof. in Appendix C. □

Equation 7 demonstrates how incorporating additional data can help reduce the bias in RDD analysis. Traditional RDD analysis of interventions typically uses only solely on a cross-section of post-intervention data. However, by incorporating pre-intervention data into the analysis, bias can be substantially reduced and, under specific conditions, eliminated. Even in cases where the bias is not completely eliminated, it can still be improved if the RDD estimation of pre-treatment data exhibits bias in the same direction as that of post-treatment RDD and if its magnitude is not too large to overcome the original bias.

3.3.1 Bias Correction

Using the MSE-optimal bandwidth for point inference can result in bandwidths that are "too large", which may seem attractive for reducing variance but can introduce first-order asymptotic bias[2]. To address this issue, we employ robust bias-corrected confidence intervals (CIs) proposed by Calonico et al. (2014). These intervals take into account the asymptotic bias of the point estimate by 1) estimating the bias and recentering the CI, and 2) incorporating the additional variance from estimating the bias for bias correction into the CI. This process requires estimating a separate local polynomial of order q, with $q > p \geq \nu$. The bias-corrected estimator is defined as:

$$\begin{aligned}
\hat{\tau}_{\nu,p,q}^{bc}(h_n, b_n) =& \tau_{\nu,p}(h_n) - h_n^{p+1-\nu}\hat{\mathbf{B}}_{\nu,p,q}(h_n, b_n), \\
\hat{\mathbf{B}}_{\nu,p,q,s}(h_n, b_n) =& \frac{\Delta\hat{\mu}_{+,q}^{(p+1)}(b_n)}{(p+1)!}\mathcal{B}_{+,\nu,p,p+1} - \frac{(-1)^{\nu+p+s}\Delta\hat{\mu}_{-,q}^{(p+1)}(b_n)}{(p+1)!}\mathcal{B}_{-,\nu,p,p+1}
\end{aligned}$$

with $\Delta\hat{\mu}_{+,q}^{(p+1)}(b_n)$ and $\Delta\hat{\mu}_{-,q}^{(p+1)}(b_n)$ being local polynomial estimations as described in Appendix A. $\hat{\mathbf{B}}_{\nu,p,q}(h_n, b_n)$ is the estimationg we get of the bias from the qth-order local polynomial.

To determine the MSE-optimal bandwidth for estimating the bias, we need to conduct a separate local polynomial estimation of order q with $q > p \geq \nu$. Once again, we aim to minimize the MSE. Following Lemma 1 and applying it to the bias estimate, we can find the MSE-optimal bandwidth for the bias estimation.

$$b_n = h_n^{MSE}{}_{p+1,q,\nu+p+1} = \left(\frac{(3+2p)\,\mathbf{V}_{p+1,q}}{n2\,(q-p)\,\mathbf{B}_{p+1,q,q+1,\nu+p+1}^2}\right)^{\frac{1}{2q+3}}$$

3.3.2 Robust Confidence Interval

Following Calonico et al. (2014), we derive a large-sample distributional approximation that accounts for the added variability introduced by the bias estimate. The large-sample approximation for the standardized t-statistic is, if $S \geq q+1$, $n\min\{h_n^{2p+3}, b_n^{2p+3}\} \times \max\{h_n^2, b_n^{2(q-p)}\} \to \infty$, then

$$T_{\nu,p,q}^{rbc}(h_n, b_n) = \frac{\hat{\tau}_{\nu,p,q}^{bc}(h_n, b_n) - \tau_\nu}{\sqrt{\mathbf{V}_{\nu,p,q}^{bc}(h_n, b_n)}} \xrightarrow{d} \mathcal{N}(0,1)$$

[2] Calonico et al. (2014)

where $\mathbf{V}^{bc}_{\nu,p,q}(h_n, b_n)$ is described in the appendix C. This justifies the following confidence intervals:

$$CI_{\nu,p,q}(h_n, b_n) = \left[\hat{\tau}^{bc}_{\nu,p,q}(h_n, b_n) \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\mathbf{V}^{bc}_{\nu,p,q}(h_n, b_n)}\right]$$

These confidence intervals have better properties compared to conventional bias-corrected intervals. They are more robust to bandwidth selection, feature coverage error decays at a faster rate, and offer shorter interval lengths, as explained by Calonico et al. (2014).

4 Validity Test

4.1 Testing if Confounding Effect is Time-Invariant

As highlighted in Section 2, the importance of Assumption 3 - that the confounding effect is constant over time - cannot be overstated. Without this assumption, all estimations become inherently flawed, resulting in biased treatment effect estimates. Therefore, an initial step when considering the applicability of a differences-in-discontinuity approach is to assess the validity of this assumption.

To accomplish this, we propose a simple test to check violations of this assumption. In essence, it involves using the available pre-treatment periods to conduct an estimation of stacked RDDs. The goal is to examine whether the RDD coefficients remain consistent across multiple periods. This involves a set of k periods where it is known that no changes occurred at the threshold.

It is crucial to note that this estimation method is not suitable if any alterations occurred at the threshold between the initial period in this sample and the period just preceding the introduction of the treatment of interest. To ensure the validity of this approach, it is necessary to select a period during which no events occurred at this threshold that could influence the observed outcome.

The procedure involves considering the following stacked RDDs regression model:

$$Y_i = \alpha_i + \sum_{k=1}^{K}\left[\beta_{-k}(Z_i - z_0) + \theta_{-k}D(Z_i - z_0)\right]T_{-k} + \varepsilon_i \tag{8}$$

where T_{-k} is a dummy indicating the period to which the data belong, that is, the RDD of which period. The hypothesis to be tested is:

$$H_0: \quad \theta_0 = \theta_{-1} = ... = \theta_{-K}$$

A Wald test is sufficient to test these hypotheses. Rejection of the null hypothesis indicates that the differences-in-discontinuity design may not be suitable for estimating the effect of treatment in this setting.

A question arises about concerning bandwidth to use when conducting this test. There are several options available, and Appendix D.1 provides details on the simulations used to assess the better bandwidth. In summary, in practice, any bandwidth derived from the data that is optimal for the specific RDD should work well.

4.2 Testing the Time-Invariance of Data-Generating Functional Form

In Section 3.3, we briefly discussed an important aspect of the DiDC method: its potential to achieve zero asymptotic bias when choosing DiDC over the standard RDD, provided that the functional forms of the data-generating processes for both groups are time-invariant. Essentially, this would mean that the derivatives of the functions on each side of the threshold at each $t = \{0, 1\}$ ($\mu_{+,1}(z)$ and $\mu_{+,0}(z)$, $\mu_{-,1}(z)$ and $\mu_{-,0}(z)$) are equal at each point of the running variable z, resulting in zero bias in the local polynomial estimation.

Claim 3. If the data-generating processes for both groups (above and below the threshold) are time-invariant, then $\Delta\mu^{(\nu)}_{+}$ and $\Delta\mu^{(\nu)}_{-}$ will be zero.

When $\Delta\mu^{(\nu)}_{+} = \Delta\mu^{(\nu)}_{-} = 0$, the bias term $\mathbf{B}_{\nu,p,r}(h_n) = \frac{\Delta\mu^{(r)}_{+}\mathcal{B}_{+,\nu,p,r}(h_n) - \Delta\mu^{(r)}_{-}\mathcal{B}_{-,\nu,p,r}(h_n)}{r!}$ is also zero. This condition implies that $\Delta\mu(z)$ is a constant, allowing unbiased point estimates through linear estimations on both sides of the threshold. As a result, two simple OLS regressions would be sufficient for accurately estimating the treatment effect.

Therefore, it can be very beneficial for researchers to know if the shapes of the data-generating functions remain stable over time. This knowledge would allow them to determine if they can use simpler and less biased estimation methods. To test this, we propose comparing the derivatives of the estimated functions on each side of the threshold. We can estimate $\mu_{+,t}(z)$ and $\mu_{-,t}(z)$ nonparametrically, for $t = \{0, 1\}$. This is, we locally approximate $\mu_t(z)$ by a polynomial of order p for z in the neighborhood of $\bar{z}$:

$$\mu_t(z) \sim \mu_t(\bar{z}) + \mu_t^{(1)}(\bar{z})\,(z-\bar{z}) + ... + \frac{\mu_t^{(p)}(\bar{z})}{p!}\,(z-\bar{z})^p$$

We perform a local polynomial regression fit for each side of the threshold and each period $t = 0, 1$:

$$\hat{\beta}_{+,t} = arg \min_{\beta \in \mathbf{R}^{p+1}} \sum_{i=1}^{n} \mathbb{1}(Z_i \geq 0)(Y_{i,t} - r_p(Z_i)'\beta)^2 K_{h_n}(Z_i - \bar{z})$$
$$\hat{\beta}_{-,t} = arg \min_{\beta \in \mathbf{R}^{p+1}} \sum_{i=1}^{n} \mathbb{1}(Z_i < 0)(Y_{i,t} - r_p(Z_i)'\beta)^2 K_{h_n}(Z_i - \bar{z})$$

where $\beta_{+,p,t} = \begin{bmatrix} \mu_{+,t} & \mu_{+,t}^{(1)} & \dots & \mu_{+,t}^{(p)}/p! \end{bmatrix}'$, and $\beta_{-,p,t} = \begin{bmatrix} \mu_{-,t} & \mu_{-,t}^{(1)} & \dots & \mu_{-,t}^{(p)}/p! \end{bmatrix}'$. The estimation procedure is similar to the one described in Section 3. We are examining whether the derivatives of the functions, rather than the levels, are consistent between periods $t = 0$ and $t = 1$ on each side. To do this, we create a test statistic inspired by the "two-sample" Kolmogorov-Smirnov test:

$$D_+ = \sup(\mathbf{1}_{p+1} - e_1)\,||\beta_{+,p,1} - \beta_{+,p,0}||$$
$$D_- = \sup(\mathbf{1}_{p+1} - e_1)\,||\beta_{-,p,1} - \beta_{-,p,0}||$$

The critical values of the test statistic depend on the choice of p, so determining the optimal p is on the agenda.

The critical values of the test statistic depend on the choice of p. Selecting the right value of p can significantly affect the outcome of the test and the conclusions drawn from the results. Therefore, determining the optimal p is on the agenda.

5 Monte Carlo Simulations

We now demonstrate the finite-sample properties of the estimator through Monte Carlo simulations and compare the performance of the proposed estimator to that of the local linear RDD estimator proposed by Calonico et al. (2014) and the nonparametric DiD regression estimator proposed by Sant'Anna and Zhao (2020). We consider Data Generating Processes (DGP) based on model 3 from Calonico et al. (2014), with small modifications that will be described.

We conduct our simulation studies in four distinct settings: with time-invariant confounding factors at the threshold, and without any confounding factors, which is the typical scenario for RDDs, for both time-invariant DGPs and time-changing DGPs. For each simulation, we conduct 1000 replications and for each replication, we consider a sample size $n = 500$, with $Z_i \sim (2\mathcal{B}(2, 4) - 1)$ where $\mathcal{B}(p_1, p_2)$ is a beta distribution with parameters p_1 and p_2. We also consider $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon = 0.1295$ and the outome generated is $Y_{i,t} = \mu_{it}(Z_i) + \varepsilon_{i,t}$. The detailed specifications and further functional forms of both models can be found in the Appendix D for added clarity. In both scenarios, we observe that our DiDC estimator performs better than the RDD estimator, with a smaller bias and improved coverage. Additionally, we find that the DiDC estimator has a similar bias and improved coverage when compared to the DiD estimator if the DGPs change over time.

5.1 Identical DGPs over time

In the first simulation, we mimic a scenario where one or more confounding discontinuities are present at the threshold $z_0 = 0$ but the DGPs are the same in both periods. This setting is comparable to that of Grembi et al. (2016), where the treatment of interest was introduced at some point between $t = 0$ and $t = 1$ and was given to units whose running variable values Z_i are above the threshold $z_0 = 0$ however there were other pre-existing treatments determined by the same threshold $z_0 = 0$ on the same running variable.

The second model we consider is a scenario with identical DGPs over time, the only distinction being that in period $t = 1$, there is an effect of treatment τ for units with $Z_i \geq 0$. Importantly, there are no other sources of discontinuity in the outcome at the threshold $z_0 = 0$, making this what would be considered an ideal scenario for estimating the effect of the treatment with a regular RDD estimation.

We present the results of these simulation studies in Table 1 and Table 2. We estimate the DiD along with the RDD and DiD and compare the average bias, median bias, root-mean-squared errors, 95% coverage probability, and 95% confidence interval length for each estimator when the treatment effect τ is equal to 1.

Figure 1: DGPs for models 1 and 2

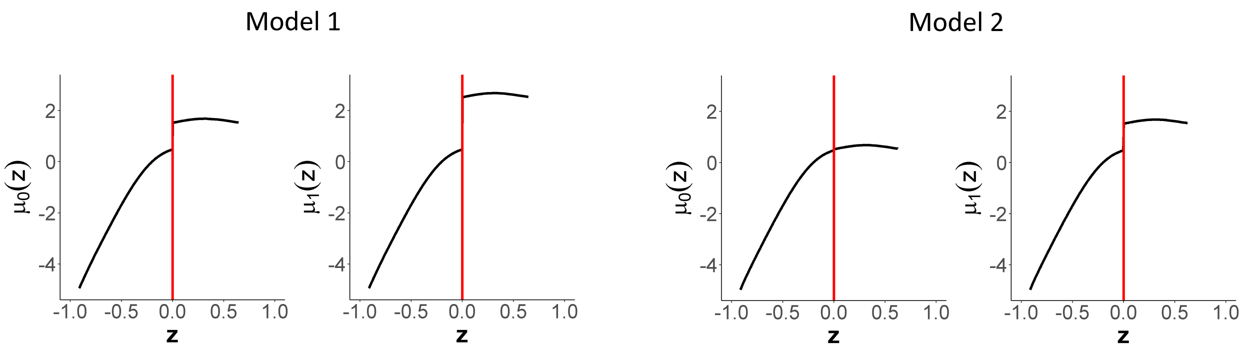


Table 1: Monte Carlo Simulations - Identical DGPs - Yes Confounder

	Av. Bias	RMSE	Coverage	CIL	Av. SE	Bandwiths h_n	b_n
RDD (Robust)	1.042	1.042	0	0.315	0.080	0.183	0.328
DiDC (RDD of Δs)	0.001	0.061	0.928	0.278	0.070	0.215	0.348
DiDC (Δ of RDDs)	0.001	0.068	0.924	0.301	0.077	0.183	0.328
DiD	0.001	0.017	1	0.559	0.143	-	-
OLS of Δs	0.001	0.023	1	0.144	0.258	-	-

Note: Simulations based on 10,000 Monte Carlo experiments with a sample size $n = 500$. RDD is the non-bias corrected RD estimator from CCT (2014), DiD is the outcome regression DiD estimator from Sant'Anna and Zhao (2020), DiDC is the estimator proposed in this paper. "Av. Bias", "RMSE", "Cover", "CIL' and "Av. SE, stand for the average simulated bias, simulated root mean-squared errors, 95% coverage probability, 95% confidence interval length and average standard error, respectively. "Bandwidths" h_n and b_n report the plug-in bandwidths for point and bias estimation, respectively.

Table 1 shows that there is a significant improvement over the RDD estimation when there is a confounding factor at the threshold. The reason for this is that the RDD method relies solely on post-treatment data and fails to account for pre-existing confounding factors, making it unsuitable for this setting. In comparison to the DiD, our estimator has closely the same bias, but the coverage seems to be more desirable since it is more informative. While the coverage of DiD is 1, meaning it covers the entire interval of the data, the coverage of the DiDC for the 95% interval is very close to 95%, at 92.8%.

Table 2: Monte Carlo Simulations - Identical DGPs - No Confounder

	Av. Bias	RMSE	Coverage	CIL	Av. SE	Bandwiths h_n	b_n
RDD (Robust)	0.044	0.073	0.903	0.314	0.080	0.183	0.328
DiDC (RDD of Δs)	0.001	0.060	0.924	0.278	0.070	0.215	0.348
DiDC (Δ of RDDs)	0.001	0.067	0.919	0.301	0.077	0.183	0.328
DiD	0.001	0.017	1	0.559	0.143	-	-
OLS of Δs	0.001	0.023	1	0.144	0.258	-	-

Note: Simulations based on 10,000 Monte Carlo experiments with a sample size $n = 500$. RDD is the non-bias corrected RD estimator from CCT (2014), DiD is the outcome regression DiD estimator from Sant'Anna and Zhao (2020), DiDC is the estimator proposed in this paper. "Av. Bias", "RMSE", "Cover", "CIL' and "Av. SE, stand for the average simulated bias, simulated root mean-squared errors, 95% coverage probability, 95% confidence interval length and average standard error, respectively. "Bandwidths" h_n and b_n report the plug-in bandwidths for point and bias estimation, respectively.

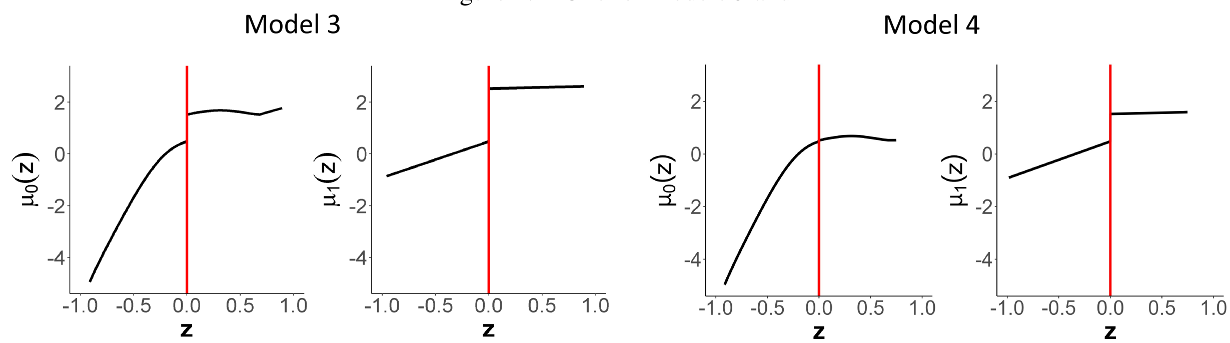
Table 2 reveals that once again our estimator has a similar bias to that of the DiD, but with more informative coverage. However, the key takeaway is that this table provides a valuable comparison between the DiDC and the RDD. In this simulation, as there were no other discontinuities at the threshold besides the treatment assignment probabilities, the scenario would be considered for the traditional use of RDD.

The results show that difference-in-discontinuities estimation has an apparent improvement over the RDD, yielding a smaller bias and better coverage. This is noteworthy as it highlights that even in cases where the RDD would traditionally be regarded as suitable, the differences-in-discontinuity approach can give more desirable results by incorporating more data into the estimation. The results hold when the functional format exhibits little temporal variation, as shown in the next section, with the additional period contributing to more reliable estimations compared to RDD, whenever the bias from the extra period $t = 0$ does not surpass the one from period $t = 1$.

5.2 Time-varying DGPs

Next, we introduce scenarios where the Data Generating Processes (DGPs) change between periods, contrasting to the prior section where DGPs were identical across time. This setup allows us to assess the performance of the difference-in-discontinuities in scenarios with small temporal DGP variations.

Figure 2: DGPs for models 3 and 4



We again do a version with confounder and a version without confounder, but now we alter the model for the period $t = 1$ so that it is a linear model derived from the original. The DGP for time $t = 0$ is identical to that of models 1 and 2. Results are shown in Table 3.

Table 3: Monte Carlo Simulations - Different DGPs

	Av. Bias	RMSE	Coverage	CIL	Av. SE	Bandwiths h_n	b_n
Yes confounder							
RDD (Robust)	1.040	1.040	0	0.289	0.074	0.215	0.348
DiDC (RDD of Δs)	0.016	0.067	0.906	0.290	0.074	0.195	0.337
DiDC (Δ of RDDs)	0.014	0.066	0.914	0.290	0.074	0.214	0.348
DiD	-1.568	1.568	0	0.384	0.098	-	-
OLS of Δs	0.684	0.684	0	0.244	0.363	-	-
No confounder							
RDD (Robust)	0.040	0.069	0.899	0.291	0.074	0.215	0.348
DiDC (RDD of Δs)	0.016	0.067	0.905	0.290	0.074	0.195	0.336
DiDC (Δ of RDDs)	0.014	0.066	0.914	0.290	0.074	0.214	0.348
DiD	-1.568	1.568	0	0.384	0.098	-	-
OLS of Δs	0.684	0.684	0	0.244	0.363	-	-

Note: Simulations based on 10,000 Monte Carlo experiments with a sample size $n = 500$. RDD is the non-bias corrected RD estimator from CCT (2014), DiD is the outcome regression DiD estimator from Sant'Anna and Zhao (2020), DiDC is the estimator proposed in this paper. "Av. Bias", "RMSE", "Cover", "CIL' and "Av. SE, stand for the average simulated bias, simulated root mean-squared errors, 95% coverage probability, 95% confidence interval length and average standard error, respectively. "Bandwidths" h_n and b_n report the plug-in bandwidths for point and bias estimation, respectively.

This table shows that in the scenario without confounding factors and time-varying DGPs, the DiD cannot be used due to the lack of parallel trends. However, the RDD can be employed, but it shows a higher bias compared to the differences-in-discontinuity approach because the GDPs, although different, are similar enough.

With confounding factors, both the RDD and DiD estimators are unsuitable. RDD remains problematic due to the confounding effects as it does not account for pre-existing discontinuities. DiD is also unsuitable because no parallel trends exist in this context. The differences-in-discontinuities method is the only approach that correctly estimates the treatment effect.

6 Empirical Illustration

We demonstrate the application of our estimator by revisiting Grembi et al. (2016), which uses data from Italian municipalities to analyze the impact of fiscal rules on municipal deficit by exploiting a government fiscal rule relaxation as a natural experiment. Our objective is to implement the DiDC approach and compare our results with the original findings. Additionally, we conduct the validity test from Section 4.1 to assess the suitability of using DiDC in this context.

The dataset comprises data from Italian municipalities, focusing on the period surrounding the government's relaxation of fiscal rules in 2001. Municipalities with fewer than 5,000 inhabitants experienced a relaxation of fiscal rules and are defined as treatment, while those with more than 5,000 inhabitants served as controls.

We reframe the data into a more manageable 2x2 design, simplifying the estimation. This transformation involves computing the differences in outcomes before and after the policy change. Grembi et al. (2016)'s original study utilized a large panel dataset, a rectangular kernel and a polynomial of degree one as in the model below:

$$Y_{it} = \delta_0 + \delta_1(X_i - c) + S_i(\gamma_0 + \gamma_1(X_i - c)) \qquad T_i[\alpha_0 + \alpha_1(X_i - c) + S_i(\beta_0 + \beta_1(X_i - c))] + \varepsilon_{it}.$$

where S_i is a dummy variable for cities below 5,000 (treatment indicator), T_i is a dummy variable for the post-treatment period and β_0 is the parameter of interest.

For our implementation, we employ the methodology described in Section 3. Table 6 compares the results from both our DiDC methods (RDD of the differences and the differences of RDDs) to Grembi et al. (2016)'s estimations. It also reports bandwidths for point estimation and number of observations.

Table 4: Effects of relaxing a fiscal rule - CCT (2014) Bandwidths

Estimators	Deficit	Fiscal Gap	Taxes
Diif-in-Disc (Grembi et al, 2016)	17.495	56.468	-76.083
	(7.737)	(32.079)	(32.597)
Bandwidth	600	513	378
Observations	2414	2136	1536
Diff-in-Disc (RDD of Δ)	23.100	44.508	-21.645
	(12.763)	(16.414)	(11.690)
Bandwidth	619	600	420
Observations	1245	1028	1030
Diff-in-Disc (Δ of RDDs)	26.864	56.043	-18.876
	(4.971)	(10.850)	(10.850)
Bandwidth	995	809	819
Observations	1456	1161	1164

We conduct the test from Section 4.1 to evaluate the assumption of time-invariant confounding effects. We use data from 1997 to 2000 to perform "stacked" RDD regressions. The results are reported in Table 6.

The hypothesis test for $Z : T_{1997} : D_{1997} = Z : T_{1998} : D_{1998} = Z : T_{1999} : D_{1999} = Z : T_{2000} : D_{2000}$ indicates a rejection of the null hypothesis, suggesting that the coefficients are not constant over time, as reported in Table 6:

In summary, our estimated treatment effects are similar in magnitude to the difference-in-discontinuities approach utilized in Grembi et al. (2016) study. Notably, our difference of RDDs approach yields smaller confidence intervals, suggesting it to be the most powerful approach. However, our validity test reveals that confounding effects vary over time, so the use of DiDC in this setting appears to be leading to biased estimates. We also replicate our estimates using Ludwig and Miller (2007)'s bandwidths with similar results. These estimates are shown in Table E in Appendix E.

7 Conclusion

The difference-in-discontinuities (DiDC) design is emerging as a promising method for estimating causal inference, addressing the limitations of both regression discontinuity (RDD) and difference-in-difference (DiD) approaches. This paper lays the theoretical groundwork for DiDC, examines its identification assumptions, estimation procedures, and asymptotic properties. We showcase its advantages through Monte Carlo simulations and an empirical application.

DiDC can handle scenarios where control and treatment groups differ significantly, violating the parallel trends assumption of DiD, or when RDD encounters confounding factors at the threshold. By incorporating more information, DiDC eliminates bias in RDD estimates under specific assumptions about the data-generating processes.

However, it is important to give the necessary importance to the identification assumptions, particularly the time-invariance of confounding effects. We propose a test based on stacked RDDs to assess its validity in practic. Additionally, DiDC requires the treatment effect to be independent of confounding policy, though we introduce a possible relaxation for potential interaction effects.

We find that the DiDC method can eliminate bias completely if the shape of the functions remains stable over time at each side of the threshold. This suggests that it could offer significant advantages over standard RDD estimations, even

Table 5:

	Dependent variable:
	imposte
Z:T_{1997}	−0.040
	(0.071)
Z:T_{1998}	0.265***
	(0.069)
Z:T_{1999}	0.176**
	(0.069)
Z:T_{2000}	0.054
	(0.068)
Z:T_{1997}:D_{1997}	0.280***
	(0.108)
Z:T_{1998}:D_{1998}	−0.369***
	(0.103)
Z:T_{1999}:D_{1999}	−0.213**
	(0.103)
Z:T_{2000}:D_{2000}	0.038
	(0.102)
Constant	121.487***
	(3.420)
Observations	4,810
R^2	0.059
Adjusted R^2	0.048
Residual Std. Error	2.137 (df = 707)
F Statistic	5.518*** (df = 8; 707)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6:

Statistic	N	Mean	St. Dev.	Min	Max
Res.Df	2	708.500	2.121	707	710
RSS	2	3,298.776	98.196	3,229.341	3,368.211
Df	1	3.000		3	3
Sum of Sq	1	138.871		138.871	138.871
F	1	10.134		10.134	10.134
Pr(>F)	1	0.00000		0.00000	0.00000

in settings where no other confounding variables exist at the threshold. We also propose test to compare the derivatives of estimated functions on either side of the threshold, allowing researchers to evaluate the stability of data-generating processes over time.

Monte Carlo simulations demonstrate DiDC's potential for improvement over RDD, yielding lower bias and better coverage. It can provide more desirable results by incorporating more data, especially when the functional form exhibits minimal temporal variation. Notably, DiDC is the only viable approach when confounding factors render both RDD and DiD unsuitable. The empirical application highlights the importance of the time-invariance assumption.

Future research directions include the development of alternative estimators that are robust to violations of identification assumptions, as well as the exploration of other confidence interval methods specifically tailored for the DiDC design. Overall, the DiDC method offers a valuable addition to the causal inference toolkit. It is applicable in settings where no other methods were previously available and shows potential to reduce bias in estimation in other settings.

References

A. Albright. The hidden effects of algorithmic recommendations. *Preprint. Last accessed March*, 28:2023, 2023.

F. Azuaga and B. Sampaio. Violência contra mulher: o impacto da lei maria da penha sobre o feminicídio no brasil. *Anais do 45o Encontro ANPEC*, 2017.

K. Butts. Geographic difference-in-discontinuities. *Applied Economics Letters*, pages 1–5, 2021.

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

S. Calonico, M. D. Cattaneo, M. H. Farrell, and R. Titiunik. *rdrobust: Robust Data-Driven Statistical Inference in Regression-Discontinuity Designs*, 2022. URL `https://CRAN.R-project.org/package=rdrobust`. R package version 2.1.1.

D. Card and A. B. Krueger. Minimum wages and employment: a case study of the fast-food industry in new jersey and pennsylvania: reply. *American Economic Review*, 90(5):1397–1420, 2000.

S. R. Cellini, F. Ferreira, and J. Rothstein. The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics*, 125(1):215–261, 2010.

L. E. Chicoine. Homicides in mexico and the expiration of the us federal assault weapons ban: a difference-in-discontinuities approach. *Journal of economic geography*, 17(4):825–856, 2017.

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

M. Frölich and S. Sperlich. *Impact evaluation*. Cambridge University Press, 2019.

H. Galindo-Silva, N. H. Some, and G. Tchuente. Fuzzy difference-in-discontinuities: Identification theory and application to the affordable care act. *arXiv preprint arXiv:1812.06537*, 2021.

R. E. García. Plant behavior & equal pay: The effect on female employment, capital investments, & productivity—a difference-in-discontinuities design. *Capital Investments, & Productivity—A Difference-in-Discontinuities Design (August 14, 2020)*, 2020.

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

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

G. W. Imbens and T. Lemieux. Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2): 615–635, 2008.

T. Lemieux and K. Milligan. Incentive effects of social assistance: A regression discontinuity approach. *Journal of Econometrics*, 142(2):807–828, 2008.

J. Ludwig and D. L. Miller. Does head start improve children's life chances? evidence from a regression discontinuity design. *The Quarterly journal of economics*, 122(1):159–208, 2007.

P. Pettersson-Lidbom. Does the size of the legislature affect the size of government? evidence from two natural experiments. *Journal of Public Economics*, 96(3-4):269–278, 2012.

P. H. Sant'Anna and J. Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1): 101–122, 2020.

Appendices

Appendix A Setup, assumptions and notation for estimation

We construct the local polynomial estimator following Calonico et al. (2014). For a given $\nu \leq p \in \mathbf{N}$, the general estimand of interest is $\tau_\nu = \Delta\mu_+ - \Delta\mu_-$ with $\Delta\mu_+^{(\nu)} = \nu! e_\nu' \delta_{+,p}$, $\Delta\mu_-^{(\nu)} = \nu! e_\nu' \delta_{-,p}$ being the νth-order derivatives of the pth-order local polynomial of the difference. The pth-order local polynomial estimators of the νth-order derivatives $\Delta\mu_{+,p}^{(\nu)}$ and $\Delta\mu_{-,p}^{(\nu)}$ are:

$$
\begin{aligned}
\Delta\hat{\mu}_{+,p}^{(\nu)}(h_n) &= \nu! e_\nu' \hat{\delta}_{+,p}(h_n) \\
\Delta\hat{\mu}_{-,p}^{(\nu)}(h_n) &= \nu! e_\nu' \hat{\delta}_{-,p}(h_n) \\
\hat{\delta}_{\Delta Y+,p}(h_n) &= arg \min_{\delta \in \mathbf{R}^{p+1}} \sum_{i=1}^{n} \mathbb{1}(Z_i \geq 0)(\Delta Y_i - r_p(Z_i)'\delta)^2 K_{h_n}(Z_i) \\
\hat{\delta}_{\Delta Y-,p}(h_n) &= arg \min_{\delta \in \mathbf{R}^{p+1}} \sum_{i=1}^{n} \mathbb{1}(Z_i < 0)(\Delta Y_i - r_p(Z_i)'\delta)^2 K_{h_n}(Z_i)
\end{aligned}
$$

where e_ν is a conformable $(\nu+1)$ unit vector, $K_h(u) = K(u/h)/h$, h_n is a positive bandwidth sequence, $r_p(x) = [1 \quad x \quad \ldots \quad x^p]'$, $\Delta Y = [\Delta Y_1 \quad \Delta Y_2 \quad \ldots \quad \Delta Y_n]'$. We set $\chi_n = [Z_1 \quad \ldots \quad Z_n]'$, $\varepsilon_{\Delta Y} = [\varepsilon_{\Delta Y,1} \quad \ldots \quad \varepsilon_{\Delta Y,n}]'$ with $\varepsilon_{\Delta Y,i} = \Delta Y_i - \mu_{\Delta Y}(Z_i)$, $\mu_{\Delta Y}(Z) = E(\Delta Y|Z)$ and $\sigma^2_{\Delta W,\Delta Y}(Z) = Cov(\Delta W, \Delta Y|Z)$. We also set

$$
\begin{aligned}
S_p(h) &= \left[(Z_1/h)^p \quad \ldots \quad (Z_n/h)^p\right]' \\
Z_p(h) &= \left[r_p(Z_1/h) \quad \ldots \quad r_p(Z_n/h)\right]' \\
W_+(h) &= diag\left(\mathbb{1}(Z_i \geq 0)K_h(Z_1), \ldots, \mathbb{1}(Z_i \geq 0)K_h(Z_n)\right) \\
W_-(h) &= diag\left(\mathbb{1}(Z_i < 0)K_h(Z_1), \ldots, \mathbb{1}(Z_i < 0)K_h(Z_n)\right) \\
\Sigma_{\Delta W \Delta Y} &= diag\left(\sigma^2_{\Delta W \Delta Y}(Z_i), \ldots, \sigma^2_{\Delta W \Delta Y}(Z_n)\right) \\
\Gamma_{+,p}(h) &= Z_p(h)'W_+(h)Z_p(h)/n \\
\Gamma_{-,p}(h) &= Z_p(h)'W_-(h)Z_p(h)/n \\
\vartheta_{+,p,q}(h) &= Z_p(h)'W_+(h)S_q(h)/n \\
\vartheta_{-,p,q}(h) &= Z_p(h)'W_-(h)S_q(h)/n \\
\Psi_{\Delta W \Delta Y+,p,q}(h,b) &= Z_p(h)'W_+(h)\Sigma_{\Delta W \Delta Y}W_+(b)Z_q(b)/n \\
\Psi_{\Delta W \Delta Y-,p,q}(h,b) &= Z_p(h)'W_-(h)\Sigma_{\Delta W \Delta Y}W_-(b)Z_q(b)/n
\end{aligned}
$$

It follows that with $H_p(h) = diag\left(1, h^{-1}, \ldots, h^{-p}\right)$:

$$
\begin{aligned}
\hat{\delta}_{\Delta Y+,p}(h_n) &= H_p(h_n)\Gamma_{+,p}^{-1}(h_n)Z_p(h_n)'W_+(h_n)\Delta Y/n \\
\hat{\delta}_{\Delta Y-,p}(h_n) &= H_p(h_n)\Gamma_{-,p}^{-1}(h_n)Z_p(h_n)'W_-(h_n)\Delta Y/n
\end{aligned}
$$

The estimand and estimators are

$$
\begin{aligned}
&\tau_\nu^{DiDC} = \Delta\mu_+^{(\nu)} - \Delta\mu_-^{(\nu)}, \quad \Delta\mu_+^{(\nu)} = \nu! e_\nu' \delta_{+,p}, \quad \Delta\mu_-^{(\nu)} = \nu! e_\nu' \delta_{-,p}, \\
&\hat{\tau}_{\nu,p}^{DiDC}(h_n) = \Delta\hat{\mu}_{+,p}^{(\nu)}(h_n) - \Delta\hat{\mu}_{-,p}^{(\nu)}(h_n), \\
&\Delta\hat{\mu}_{+,p}^{(\nu)}(h_n) = \nu! e_\nu' \hat{\delta}_{+,p}(h_n), \quad \Delta\hat{\mu}_{-,p}^{(\nu)}(h_n) = \nu! e_\nu' \hat{\delta}_{-,p}(h_n),
\end{aligned}
$$

where, for any random variables W and X, and $s \in \mathbb{N}$,

$$\begin{aligned}
&\Delta\mu_{X+}^{(s)} = \lim_{x\to 0^+} \frac{\partial^s}{\partial z^s}\Delta\mu_X(z), \quad \Delta\mu_{X-}^{(s)} = \lim_{x\to 0^-} \frac{\partial^s}{\partial z^s}\Delta\mu_X(z),\\
&\Delta\mu_X(z) = \mathbb{E}[X \mid Z = z],\\
&\sigma_{X+}^2 = \lim_{z\to 0^+} \sigma_X^2(z), \quad \sigma_{X-}^2 = \lim_{z\to 0^-} \sigma_X^2(z),\\
&\sigma_X^2(z) = \mathbb{V}[X \mid Z = z],\\
&\sigma_{WX+}^2 = \lim_{z\to 0^+} \sigma_{WX}^2(z), \quad \sigma_{WX_-}^2 = \lim_{z\to 0^-} \sigma_{WX}^2(z),\\
&\sigma_{WX}^2(z) = \mathbb{C}[W, X \mid X = z].
\end{aligned}$$

We employ the following assumptions on the *sharp* model for the non-parametric local polynomial regression estimation:

Assumption A.1. *For some $\mathcal{K}_0 > 0$, the following holds in the neighborhood $(-\mathcal{K}_0, \mathcal{K}_0)$ around the cutoff $z_0 = 0$:*

(a) $E\left[\Delta Y_i^4 | Z_i = z\right]$ *is bounded, and the density $f(z)$ of the random sample Z_i is continuous and bounded away from zero.*

(b) $\Delta\mu_-(z) = E\left[\Delta Y_i(0)|Z_i = z\right]$ *and* $\Delta\mu_+(z) = E\left[\Delta Y_i(1)|Z_i = z\right]$ *are S times continuously differentiable.*

(c) $\sigma_-^2(z) = V\left[\Delta Y_i(0)|Z_i = z\right]$ *and* $\sigma_+^2(z) = V\left[\Delta Y_i(1)|Z_i = z\right]$ *are continuous and bounded away from zero.*

We also impose the following assumption on the kernel function to be employed in the estimator. This assumption allows for most of the commonly used kernels.

Assumption A.2. *For some $\mathcal{K} > 0$, the kernel function $k(\cdot) : [0, \mathcal{K}] \to \mathbb{R}$ is bounded and nonnegative, zero outside its support, symmetric around z_0 and positive and continuous on $(0, \mathcal{K})$.*

Assumptions A.1 and A.2 limit the behavior of $E\left(\Delta Y_i | Z_i = z_0\right)$ in the vicinity of the cutoff $z_0 = 0$.

Appendix B Preliminary lemmas and results

Before proceeding, please refer to Appendix A for notation. This appendix restates, with minor adaptations, several lemmas, results and proofs from Calonico et al. (2014) that are necessary for deriving the asymptotic results.

The following lemma establishes convergence in probability of the sample matrices $\Gamma_{-,p}(h_n)$, $\boldsymbol{\vartheta}_{-,p,q}(h_n)$, $\Psi_{-,p}(h_n)$ and $\Gamma_{+,p}(h_n)$, $\boldsymbol{\vartheta}_{+,p,q}(h_n)$, $\Psi_{+,p}(h_n)$ to their expectation counterparts, and characterizes those limits.

Lemma B.1. Suppose Assumptions $1-2$ hold, and $nh_n \to \infty$.

(a) If $\kappa h_n < \kappa_0$, then:

(a.1) $\Gamma_{+,p}(h_n) = \tilde{\Gamma}_p(h_n) + o_p(1)$ with $\tilde{\Gamma}_{+,p}(h_n) = \int_0^\infty K(u) r_p(u) r_p(u)' f(uh_n)\,\mathrm{d}u \asymp \Gamma_p$

(a.2) $\Gamma_{-,p}(h_n) = H_p(-1)\tilde{\Gamma}_p(h_n) H_p(-1) + o_p(1)$ with $\tilde{\Gamma}_{-,p}(h_n) = \int_0^\infty K(u) r_p(u) r_p(u)' f(-uh_n)\,\mathrm{d}u \asymp \Gamma_p$,

(a.3) $\boldsymbol{\vartheta}_{+,p,q}(h_n) = \tilde{\boldsymbol{\vartheta}}_{+,p,q}(h_n) + o_p(1)$ with $\tilde{\boldsymbol{\vartheta}}_{+,p,q}(h_n) = \int_0^\infty K(u) r_p(u) u^q f(uh_n)\,\mathrm{d}u \asymp \boldsymbol{\vartheta}_{p,q}$,

(a.4) $\boldsymbol{\vartheta}_{-,p,q}(h_n) = (-1)^q H_p(-1)\tilde{\boldsymbol{\vartheta}}_{-,p,q}(h_n) + o_p(1)$ with $\tilde{\boldsymbol{\vartheta}}_{-,p,q}(h_n) = \int_0^\infty K(u) r_p(u) u^q f(-uh_n)\,\mathrm{d}u \asymp \boldsymbol{\vartheta}_{p,q}$,

(a.5) $h_n \Psi_{+,p}(h_n) = \tilde{\Psi}_{+,p}(h_n) + o_p(1)$ with $\tilde{\Psi}_{+,p}(h_n) = \int_0^\infty K(u)^2 r_p(u) r_p(u)' \boldsymbol{\sigma}_+^2(uh_n) f(uh_n)\,\mathrm{d}u \asymp \Psi_p$,

(a.6) $h_n \Psi_{-,p}(h_n) = H_p(-1)\tilde{\Psi}_{-,p}(h_n) H_p(-1) + o_p(1)$ with $\tilde{\Psi}_{-,p}(h_n) = \int_0^\infty K(u)^2 r_p(u) r_p(u)' \sigma_-^2(-uh_n) f(-uh_n)\,\mathrm{d}u \asymp \Psi_p$.

(b) If $h_n \to 0$, then

(b.1) $\tilde{\Gamma}_{+,p}(h_n) = f\Gamma_p + o(1)$ and $\tilde{\Gamma}_{-,p}(h_n) = f\Gamma_p + o(1)$,

(b.2) $\tilde{\vartheta}_{+,p,q}(h_n) = f\vartheta_{p,q} + o(1)$ and $\tilde{\vartheta}_{-,p,q}(h_n) = f\boldsymbol{\vartheta}_{p,q} + o(1)$,

(b.3) $\tilde{\Psi}_{+,p}^{+,p,q}(h_n) = \sigma_+^2 f\Psi_p + o(1)$ and $\tilde{\Psi}_{-,p}(h_n) = \sigma_-^2 f\Psi_p + o(1)$.

Proof. For part (a.5), change of variable implies

$$\begin{aligned}
\mathbb{E}\left[\Psi_{+,p}(h_n)\right] &= \mathbb{E}\left[h_n Z_p(h)' W_+(h) \Sigma_{\Delta W \Delta Y} W_+(b) Z_q(b)/n\right] \\
&= \int_0^\infty K\left(\frac{z}{h_n}\right)^2 r_p\left(\frac{z}{h_n}\right) r_p\left(\frac{z}{h_n}\right)' \sigma_+^2 f(z)\mathrm{d}z \\
&= \int_0^\infty K(u)^2 r_p(u) r_p(u)' \sigma_+^2(uh_n) f(uh_n)\,\mathrm{d}u \\
&= \tilde{\Psi}_{+,p}(h_n),
\end{aligned}$$

and $h_n^2 \mathbb{E}\left[\left|\Psi_{+,p}(h_n) - \mathbb{E}\left[\Psi_{+,p}(h_n)\right]\right|^2\right] = n^{-1}h_n^{-1}\int_0^\infty K(u)^4 \left|r_p(u)\right|^4 f(uh_n)\,\mathrm{d}u = O\left(n^{-1}h_n^{-1}\right)$, provided $\kappa h_n < \kappa_0$. For part (a.6),

$$\begin{aligned}
\mathbb{E}\left[h_n \Psi_{-,p}(h_n)\right] &= h_n^{-1}\int_{-\infty}^0 K(u/h_n)^2 r_p(u/h_n) r_p(u/h_n)' \sigma_-^2(u) f(u)\mathrm{d}u \\
&= H_p(-1)\tilde{\Psi}_{-,p}(h_n) H_p(-1),
\end{aligned}$$

and the rest is proven as above. Also, note that $\tilde{\Psi}_{+,p}^{+,p,q}(h_n) = \sigma_+^2 f\Psi_p + o(1)$ and $\tilde{\Psi}_{-,p}(h_n) = \sigma_-^2 f\Psi_p + o(1)$ if $h_n \to 0$, by continuity of $\sigma_+^2(u)$, $\sigma_-^2(u)$ and $f(u)$, which proves part (b.3).

Proofs for the other items follow similarly to the one above and can be found in Calonico et al. (2014), as they are identical to those provided there. □

Let $s, \ell \in \mathbb{N}$ with $s \leq \ell$. The following lemma gives the asymptotic bias, variance, and distribution for the ℓ th-order local polynomial estimator of $\Delta\mu_+^{(s)}$ and $\Delta\mu_-^{(s)}$:

$$\begin{aligned}
&\Delta\hat{\mu}_{+,\ell}^{(s)}\left(h_n\right) = s! e_s' \delta_{+,\ell}\left(h_n\right), \\
&\Delta\hat{\delta}_{+,\ell}\left(h_n\right) = H_\ell\left(h_n\right) \Gamma_{+,\ell}^{-1}\left(h_n\right) Z_\ell\left(h_n\right)' W_+\left(h_n\right) Y/n, \\
&\Delta\hat{\mu}_{-,\ell}^{(s)}\left(h_n\right) = s! e_s' \hat{\delta}_{-,\ell}\left(h_n\right), \\
&\hat{\delta}_{-,\ell}\left(h_n\right) = H_\ell\left(h_n\right) \Gamma_{-,\ell}^{-1}\left(h_n\right) Z_\ell\left(h_n\right)' W_-\left(h_n\right) Y/n.
\end{aligned}$$

Lemma B.2. Suppose Assumptions 1 and 2 hold with $S \geq \ell + 2$, and $nh_n \to \infty$.

(B) If $h_n \to 0$, then

$$\begin{aligned}
&\mathbb{E}\left[\Delta\hat{\mu}_{+,\ell}^{(s)}\left(h_n\right) \mid \mathcal{X}_n\right] = s! e_s' \delta_{+,\ell} + h_n^{1+\ell-s} \frac{\Delta\mu_+^{(\ell+1)}}{(\ell+1)!} \mathcal{B}_{+,s,\ell,\ell+1}\left(h_n\right) \\
&+ h_n^{2+\ell-s} \frac{\Delta\mu_+^{(\ell+2)}}{(\ell+2)!} \mathcal{B}_{+,s,\ell,\ell+2}\left(h_n\right) + o_p\left(h_n^{2+\ell-s}\right), \\
&\mathcal{B}_{+,s,\ell,r}\left(h_n\right) = s! e_s' \Gamma_{+,\ell}^{-1}\left(h_n\right) \vartheta_{+,\ell,r}\left(h_n\right) = s! e_s' \Gamma_\ell^{-1} \vartheta_{\ell,r} + o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\left[\Delta\hat{\mu}_{-,\ell}^{(s)}\left(h_n\right) \mid \mathcal{X}_n\right] = s! e_s' \delta_{-,\ell} + h_n^{1+\ell-s} \frac{\Delta\mu_-^{(\ell+1)}}{(\ell+1)!} \mathcal{B}_{-,s,\ell,\ell+1}\left(h_n\right) \\
&+ h_n^{2+\ell-s} \frac{\Delta\mu_-^{(\ell+2)}}{(\ell+2)!} \mathcal{B}_{-,s,\ell,\ell+2}\left(h_n\right) + o_p\left(h_n^{2+\ell-s}\right), \\
&\mathcal{B}_{-,s,\ell,r}\left(h_n\right) = s! e_s' \Gamma_{-,\ell}^{-1}\left(h_n\right) \vartheta_{-,\ell,r}\left(h_n\right) = (-1)^{s+r} s! e_s' \Gamma_\ell^{-1} \boldsymbol{\vartheta}_{\ell,r} + o_p(1).
\end{aligned}$$

(V) If $h_n \to 0$, then $\mathbb{V}\left[\Delta\hat{\mu}_{+,\ell}^{(s)}\left(h_n\right) \mid \mathcal{X}_n\right] = \mathcal{V}_{+,s,\ell}\left(h_n\right)$ with

$$\begin{aligned}
\mathcal{V}_{+,s,\ell}\left(h_n\right) &= \frac{1}{nh_n^{2s}} s!^2 e_s' \Gamma_{+,\ell}^{-1}\left(h_n\right) \Psi_{+,\ell}\left(h_n\right) \Gamma_{+,\ell}^{-1}\left(h_n\right) e_s \\
&= \frac{1}{nh_n^{1+2s}} \frac{\sigma_+^2}{f} s!^2 e_s' \Gamma_\ell^{-1} \Psi_\ell \Gamma_\ell^{-1} e_s \left[1 + o_p(1)\right],
\end{aligned}$$

and $\mathbb{V}\left[\Delta\hat{\mu}_{-,\ell}^{(s)}\left(h_n\right) \mid \mathcal{X}_n\right] = \mathcal{V}_{-,s,\ell}\left(h_n\right)$ with

$$\begin{aligned}
\mathcal{V}_{-,s,\ell}\left(h_n\right) &= \frac{1}{nh_n^{2s}} s^2 e_s' \Gamma_{-\ell}^{-1}\left(h_n\right) \Psi_{-,\ell}\left(h_n\right) \Gamma_{-,\ell}^{-1}\left(h_n\right) e_s \\
&= \frac{1}{nh_n^{1+2s}} \frac{\sigma_-^2}{f} s!^2 e_s' \Gamma_\ell^{-1} \Psi_\ell \Gamma_\ell^{-1} e_s \left[1 + o_p(1)\right].
\end{aligned}$$

(D) If $nh_n^{2\ell+5} \to 0$, then

$$\frac{\Delta\hat{\mu}_{+,\ell}^{(s)}\left(h_n\right) - \Delta\mu_+^{(s)} - h_n^{1+\ell-s} \frac{\Delta\mu_+^{(\ell+1)}}{(\ell+1)!} \mathcal{B}_{+,s,\ell,\ell+1}\left(h_n\right)}{\sqrt{\mathcal{V}_{+,s,\ell}\left(h_n\right)}} \to_d \mathcal{N}(0,1)$$

and

$$\frac{\Delta\hat{\mu}_{-,\ell}^{(s)}\left(h_n\right) - \Delta\mu_-^{(s)} - h_n^{1+\ell-s} \frac{\Delta\mu_-^{(\ell+1)}}{(\ell+1)!} \mathcal{B}_{-,s,\ell,\ell+1}\left(h_n\right)}{\sqrt{\mathcal{V}_{-,s,\ell}\left(h_n\right)}} \to_d \mathcal{N}(0,1).$$

Proof. For part (B), a Taylor series expansion gives

$$\begin{aligned}
&\mathbb{E}[s!\ \hat{\delta}_{+,\ell}\left(h_n\right) \mid \mathcal{X}_n\Big] \\
=&s!\delta_{+,\ell} + h_n^{\ell+1} H_\ell\left(h_n\right)\Gamma_{+,\ell}^{-1}\left(h_n\right) Z_\ell\left(h_n\right) W_+\left(h_n\right) S_{\ell+1}\left(h_n\right) s!\frac{\Delta\mu_+^{(\ell+1)}}{(\ell+1)!} \\
&+ h_n^{\ell+2} H_\ell\left(h_n\right)\Gamma_{+,\ell}^{-1}\left(h_n\right) Z_\ell\left(h_n\right) W_+\left(h_n\right) S_{\ell+2}\left(h_n\right) s!\frac{\Delta\mu_+^{(\ell+2)}}{(\ell+2)!} + H_\ell\left(h_n\right) o_p\left(h_n^{\ell+2}\right) \\
=&s!\delta_{+,\ell} + h_n^{\ell+1} H_\ell\left(h_n\right) s!\frac{\Delta\mu_+^{(\ell+1)}}{(\ell+1)!}\Gamma_{+,\ell}^{-1}\left(h_n\right)\vartheta_{+,\ell,\ell+1}\left(h_n\right) \\
&+ h_n^{\ell+2} H_\ell\left(h_n\right) s!\frac{\Delta\mu_+^{(\ell+2)}}{(\ell+2)!}\Gamma_{+,\ell}^{-1}\left(h_n\right)\vartheta_{+,\ell,\ell+2}\left(h_n\right) + H_\ell\left(h_n\right) o_p\left(h_n^{\ell+2}\right),
\end{aligned}$$

and the result for $\mathbb{E}\left[\Delta\hat{\mu}_{+,\ell}^{(s)}\left(h_n\right) \mid \mathcal{X}_n\right]$ follows by $e_s' H_\ell\left(h_n\right) = h_n^{-s}$ and Lemma B.1. Next, for $\mathbb{E}\left[\Delta\hat{\mu}_{-,\ell}^{(s)}\left(h_n\right) \mid \mathcal{X}_n\right]$ the same calculations apply, with only a modification for $\mathcal{B}_{-,s,\ell,r}\left(h_n\right)$ because, by Lemma B.1,

$$\begin{aligned}
\mathcal{B}_{-,s,\ell,r}\left(h_n\right) =&s! e_s'\Gamma_{-,\ell}^{-1}\left(h_n\right)\vartheta_{-,\ell,r}\left(h_n\right) \\
=&s! e_s'\left[H_\ell(-1)\tilde{\Gamma}_{-,\ell}^{-1}\left(h_n\right) H_\ell(-1)\right]\left[(-1)^r H_\ell(-1)\vartheta_{-,\ell,r}\left(h_n\right)\right] \\
&+ o_p(1) \\
=&(-1)^{s+r} s! e_s'\tilde{\Gamma}_{-,\ell}^{-1}\left(h_n\right)\vartheta_{-,\ell,r}\left(h_n\right) + o_p(1),
\end{aligned}$$

because $e_s' H_\ell(-1) = (-1)^s$ and $H_\ell(-1) H_\ell(-1) = I_{\ell+1}$. For part (V), simply note that

$$\begin{aligned}
\mathbb{V}\left[s! e_s'\hat{\delta}_{+,\ell}\left(h_n\right) \mid \mathcal{X}_n\right] =&s!^2 e_s' H_\ell\left(h_n\right)\Gamma_{+,\ell}^{-1}\left(h_n\right) Z_\ell\left(h_n\right) W_+\left(h_n\right)\Sigma \\
&\times W_+\left(h_n\right) Z_\ell\left(h_n\right)\Gamma_{+,\ell}^{-1}\left(h_n\right) H_\ell\left(h_n\right) e_s/n \\
=&h_n^{-2s} s!^2 e_s'\Gamma_{+,\ell}^{-1}\left(h_n\right) Z_\ell\left(h_n\right) W_+\left(h_n\right)\Sigma \\
&\times W_+\left(h_n\right) Z_\ell\left(h_n\right)\Gamma_{+,\ell}^{-1}\left(h_n\right) e_s/n \\
=&n^{-1} h_n^{-2s} s!^2 e_s'\Gamma_{+,\ell}^{-1}\left(h_n\right)\Psi_{+,\ell}\left(h_n\right)\Gamma_{+,\ell}^{-1}\left(h_n\right) e_s \\
=&\mathcal{V}_{+,\ell,s}\left(h_n\right),
\end{aligned}$$

and the result follows by Lemma B.1. The proof of $\mathbb{V}\left[s! e_s'\hat{\delta}_{-,\ell}\left(h_n\right) \mid \mathcal{X}_n\right]$ is analogous.

For part (D), using the previous results, we have

$$= \frac{s! e_s'\hat{\boldsymbol{\delta}}_{+,\ell}\left(h_n\right) - s! e_s'\delta_{+,\ell} - h_n^{1+\ell-s}\frac{\Delta\mu_+^{(\ell+1)}}{(\ell+1)!}\mathcal{B}_{+,s,\ell,\ell+1}\left(h_n\right)}{\sqrt{\mathcal{V}_{+,s,\ell}\left(h_n\right)}}$$

The result for $\Delta\hat{\mu}_{-,\ell}^{(s)}\left(h_n\right)$ can be established the same way. This concludes the proof. Q.E.D. ☐

Let ν, p, $q \in \mathbb{N}$ with $\nu \leq p < q$. The nex lemma dives the asymptotic bias, variance and distribution for the pth-order local polynomial estimator of $\Delta\mu_+^{(\nu)}$ and $\Delta\mu_-^{(\nu)}$ with bias correction constructed using a qth-order local polynomial.

Lemma B.3. Suppose Assumptions 1 and 2 hold with $S \geq q+1$, and $n\min\{h_n, b_n\} \to \infty$. (B) If $\max\{h_n, b_n\} \to 0$, then

$$\begin{aligned}
&\mathbb{E}\left[\Delta\hat{\mu}_{+,p,q}^{(\nu)\,\mathrm{bc}}\left(h_n, b_n\right) \mid \mathcal{X}_n\right] \\
&= \nu! e_\nu'\delta_{+,p} + h_n^{2+p-\nu}\frac{\Delta\mu_+^{(p+2)}}{(p+2)!}\mathcal{B}_{+,\nu,p,p+2}\left(h_n\right)\{1+o_p(1)\} \\
&\quad - h_n^{p+1-\nu} b_n^{q-p}\frac{\Delta\mu_+^{(q+1)}}{(q+1)!}\mathcal{B}_{+,p+1,q,q+1}\left(b_n\right)\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!}\{1+o_p(1)\}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\left[\Delta\hat{\mu}^{(\nu)\mathrm{bc}}_{-,p,q}\left(h_n, b_n\right) \mid \mathcal{X}_n\right] \\
&= \nu! e_\nu' \delta_{-,p} + h_n^{2+p-\nu} \frac{\Delta\mu_-^{(p+2)}}{(p+2)!} \mathcal{B}_{-,\nu,p,p+2}\left(h_n\right)\{1+o_p(1)\} \\
&\quad - h_n^{p+1-\nu} b_n^{q-p} \frac{\Delta\mu_-^{(q+1)}}{(q+1)!} \mathcal{B}_{-,p+1,q,q+1}\left(b_n\right) \frac{\mathcal{B}_{-,\nu,p,p+1}\left(h_n\right)}{(p+1)!}\{1+o_p(1)\}.
\end{aligned}$$

(V) If $n \min\left\{h_n, b_n\right\} \to \infty$, then $\mathbb{V}\left[\Delta\hat{\mu}^{(\nu)\mathrm{bc}}_{+,p,q}\left(h_n, b_n\right) \mid \mathcal{X}_n\right] = \mathcal{V}^{\mathrm{bc}}_{+,v,p,q}\left(h_n, b_n\right)$, where

$$\begin{aligned}
\mathcal{V}^{\mathrm{bc}}_{+,\nu,p,q}(h, b) =& \mathcal{V}_{+,\nu,p}(h) + h_n^{2(p+1-\nu)} \mathcal{V}_{+,p+1,q}(b) \frac{\mathcal{B}_{+,\nu,p,p+1}(h)^2}{(p+1)!^2} \\
& - 2h^{p+1-\nu} \mathcal{C}_{+,\nu,p,q}(h, b) \frac{\mathcal{B}_{+,v,p,p+1}(h)}{(p+1)!}, \\
\mathcal{C}_{+,\nu,p,q}(h, b) =& \frac{1}{nh^\nu b^{p+1}} \nu!(p+1)! e_\nu' \Gamma_{+,p}^{-1}(h) \Psi_{+,p,q}(h, b) \Gamma_{+,q}^{-1}(b) e_p,
\end{aligned}$$

and $\mathbb{V}\left[\Delta\hat{\mu}^{(\nu)\,\mathrm{bc}}_{-,p,q}\left(h_n, b_n\right) \mid \mathcal{X}_n\right] = \mathcal{V}^{\mathrm{bc}}_{-,\nu,p,q}\left(h_n, b_n\right)$, where

$$\begin{aligned}
\mathcal{V}^{\mathrm{bc}}_{-,\nu,p,q}(h, b) =& \mathcal{V}_{-,\nu,p}(h) + h_n^{2(p+1-\nu)} \mathcal{V}_{-,p+1,q}(b) \frac{\mathcal{B}_{-,\nu,p,p+1}(h)^2}{(p+1)!^2} \\
& - 2h^{p+1-\nu} \mathcal{C}_{-,\nu,p,q}(h, b) \frac{\mathcal{B}_{-,v,p,p+1}(h)}{(p+1)!} \\
\mathcal{C}_{-,\nu,p,q}(h, b) =& \frac{1}{nh^\nu b^{p+1}} \nu!(p+1)! e_\nu' \Gamma_{-,p}^{-1}(h) \Psi_{-,p,q}(h, b) \Gamma_{-,q}^{-1}(b) e_p
\end{aligned}$$

(D) If $n \min\left\{h_n^{2p+3}, b_n^{2p+3}\right\} \max\left\{h_n^2, b_n^{2(q-p)}\right\} \to 0$, and $\kappa \max\left\{h_n, b_n\right\} < \kappa_0$, then

$$\frac{\Delta\hat{\mu}^{(\nu)\mathrm{bc}}_{+,p,q}\left(h_n, b_n\right) - \nu! e_\nu' \delta_{+,p}}{\sqrt{\mathcal{V}^{\mathrm{bc}}_{+,\nu,p,q}\left(h_n, b_n\right)}} \to_d \mathcal{N}(0,1)$$

and

$$\frac{\Delta\hat{\mu}^{(\nu)\mathrm{bc}}_{-,p,q}\left(h_n, b_n\right) - \nu! e_\nu' \delta_{-,p}}{\sqrt{\mathcal{V}^{\mathrm{bc}}_{-,\nu,p,q}\left(h_n, b_n\right)}} \to_d \mathcal{N}(0,1)$$

Proof. For part (B), first note that $\mathbb{E}\left[\Delta\hat{\mu}^{(\nu)bc}_{+,p,q}\left(h_n, b_n\right) \mid \mathcal{X}_n\right] = B_1 - B_2$ with $B_1 = \mathbb{E}\left[\nu! e_\nu' \hat{\delta}_{+,p}\left(h_n\right) \mid \mathcal{X}_n\right]$ and $B_2 = h_n^{p+1-\nu} \mathbb{E}\left[e_{p+1}' \hat{\delta}_{+,q}\left(b_n\right) \mid \mathcal{X}_n\right] \mathcal{B}_{+,\nu,p}\left(h_n\right)$. By Lemma B.2, with $s = \nu$ and $\ell = p$, we have

$$\begin{aligned}
B_1 =& \nu! e_\nu' \delta_{+,p} + h_n^{1+p-\nu} \frac{\Delta\mu_+^{(p+1)}}{(p+1)!} \mathcal{B}_{+,v,p,p+1}\left(h_n\right) \\
& + h_n^{2+p-\nu} \frac{\Delta\mu_+^{(p+2)}}{(p+2)!} \mathcal{B}_{+,v,p,p+2}\left(h_n\right) + o_p\left(h_n^{2+p-\nu}\right)
\end{aligned}$$

Similarly, by Lemma B.2, with $s = p+1$ and $\ell = q$, we have

$$\begin{aligned}
&\mathbb{E}\left[(p+1)! e_{p+1}' \hat{\delta}_{+,q}\left(b_n\right) \mid \mathcal{X}_n\right] \\
&= (p+1)! e_{p+1}' \delta_{+,q} + b_n^{q-p} \frac{\Delta\mu_+^{(q+1)}}{(q+1)!} \mathcal{B}_{+,p+1,q,q+1}\left(b_n\right) + o_p\left(b_n^{q-p}\right)
\end{aligned}$$

and hence

$$\begin{aligned} B_2 =& h_n^{p+1-\nu}\mathbb{E}\left[(p+1)!e_{p+1}'\hat{\delta}_{+,q}\left(b_n\right) \mid \mathcal{X}_n\right]\frac{\mathcal{B}_{+,v,p,p+1}\left(h_n\right)}{(p+1)!} \\ =& h_n^{p+1-\nu}\left(e_{p+1}'\delta_{+,q)}\right)\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right) \\ &+ h_n^{p+1-\nu}b_n^{q-p}\frac{\Delta\mu_+^{(q+1)}}{(q+1)!}\mathcal{B}_{+,p+1,q,q+1}\left(b_n\right)\frac{\mathcal{B}_{+,v,p,p+1}\left(h_n\right)}{(p+1)!} \\ &+ h_n^{p+1-\nu}o_p\left(b_n^{q-p}\right)\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right) \end{aligned}$$

Collecting terms, the result in part (B) follows:

$$\begin{aligned} &\mathbb{E}\Big[\nu!e_\nu'\hat{\delta}_{+,p,q}^{\mathrm{bc}}\left(h_n,b_n\right) \mid \mathcal{X}_n\Big] \\ =&\nu!e_\nu'\delta_{+,p} + h_n^{2+p-\nu}\frac{\Delta\mu_+^{(p+2)}}{(p+2)!}\mathcal{B}_{+,\nu,p,p+2}\left(h_n\right)\{1+o_p(1)\} \\ &- h_n^{p+1-\nu}b_n^{q-p}\frac{\Delta\mu_+^{(q+1)}}{(q+1)!}\mathcal{B}_{+,p+1,q,q+1}\left(b_n\right)\frac{\mathcal{B}_{+,\nu,p}\left(h_n\right)}{(p+1)!}\{1+o_p(1)\} \end{aligned}$$

For part (V), first note that $\mathbb{V}\left[\Delta\hat{\mu}_{+,p,q}^{(\nu)\,\mathrm{bc}}\left(h_n,b_n\right) \mid \mathcal{X}_n\right] = V_1+V_2-2C_{12}$ where, using Lemma B.2 with $s=\nu$ and $\ell=p$,

$$V_1 = \mathbb{V}\left[\nu!e_\nu'\hat{\delta}_{+,p}\left(h_n\right) \mid \mathcal{X}_n\right] = \mathbb{V}\left[\Delta\hat{\mu}_{+,p}^{(\nu)}\left(h_n\right) \mid \mathcal{X}_n\right] = \mathcal{V}_{+,v,p}\left(h_n\right)$$

and, using LemmaB.2 with $s=p+1$ and $\ell=q$,

$$\begin{aligned} V_2 &= \mathbb{V}\left[h_n^{p+1-\nu}\left(e_{p+1}'\hat{\delta}_{+,q}\left(b_n\right)\right)\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right) \mid \mathcal{X}_n\right] \\ &= h_n^{2(p+1-\nu)}\mathbb{V}\left[(p+1)!e_{p+1}'\hat{\boldsymbol{\delta}}_{+,q}\left(b_n\right) \mid \mathcal{X}_n\right]\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)^2}{(p+1)!^2} \\ &= h_n^{2(p+1-\nu)}\mathcal{V}_{+,p+1,q}\left(b_n\right)\frac{\mathcal{B}_{+,v,p,p+1}\left(h_n\right)^2}{(p+1)!^2} \end{aligned}$$

and

$$\begin{aligned} C_{12} &= \mathbb{C}\left[\nu!e_\nu'\hat{\delta}_{+,p}\left(h_n\right), h_n^{p+1-\nu}\left(e_{p+1}'\hat{\delta}_{+,q}\left(b_n\right)\right)\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right) \mid \mathcal{X}_n\right] \\ &= h_n^{p+1-\nu}\mathbb{C}\left[\nu!e_\nu'\hat{\delta}_{+,p}\left(h_n\right), (p+1)!e_{p+1}'\hat{\delta}_{+,q}\left(b_n\right) \mid \mathcal{X}_n\right]\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!} \end{aligned}$$

with

$$\begin{aligned} &\mathbb{C}\Big[e_\nu'\hat{\delta}_{+,p}\left(h_n\right), e_{p+1}'\hat{\delta}_{+,q}\left(b_n\right) \mid \mathcal{X}_n\Big] \\ =&h_n^{-\nu}e_\nu'\Gamma_{+,p}^{-1}\left(h_n\right)X_p\left(h_n\right)W_+\left(h_n\right)\mathbb{C}\left[Y,Y \mid \mathcal{X}_n\right] \\ &\times W_+\left(b_n\right)X_q\left(b_n\right)\Gamma_{+,q}^{-1}\left(b_n\right)e_{p+1}b_n^{-p-1}/n^2 \\ =&\frac{1}{nh_n^v b_n^{p+1}}\nu!(p+1)!e_\nu'\Gamma_{+,\ell}^{-1}\left(h_n\right)\Psi_{+,p,q}\left(h_n,b_n\right)\Gamma_{+,q}^{-1}\left(b_n\right)e_{p+1}. \end{aligned}$$

Thus, collecting terms, we obtain the result in part (V).

Proof for the control group (subindex "-") is analogous. □

Lemma B.4. Suppose Assumptions $1-2$ hold with $S \geq p+2$, and $nh_n \to \infty$. Let $r \in \mathbb{N}$.

(B) If $h_n \to 0$, then

$$\begin{aligned} \mathbb{E}\left[\hat{\tau}_{\nu,p}\left(h_n\right) \mid \mathcal{X}_n\right] =& \tau_\nu + h_n^{p+1-\nu}\mathbf{B}_{\nu,p,p+1}\left(h_n\right) \\ &+ h_n^{p+2-\nu}\mathbf{B}_{\nu,p,p+2}\left(h_n\right) + o_p\left(h_n^{p+2-\nu}\right) \end{aligned}$$

where

$$\mathbf{B}_{\nu,p,r}(h_n) = \frac{\Delta\mu_+^{(r)}}{r!}\mathcal{B}_{+,\nu,p,r}(h_n) - \frac{\Delta\mu_-^{(r)}}{r!}\mathcal{B}_{-,\nu,p,r}(h_n)$$
$$\mathcal{B}_{+,v,p,r}(h_n) = \nu! e_\nu' \Gamma_{+,p}^{-1}(h_n)\,\vartheta_{+,p,r}(h_n) = \nu! e_\nu' \Gamma_p^{-1}\boldsymbol{\vartheta}_{p,r} + o_p(1)$$
$$\mathcal{B}_{-,\nu,p,r}(h_n) = \nu! e_\nu' \Gamma_{-,p}^{-1}(h_n)\,\vartheta_{-,p,r}(h_n) = (-1)^{\nu+r}\nu! e_\nu' \Gamma_p^{-1}\boldsymbol{\vartheta}_{p,r} + o_p(1)$$

(V) If $h_n \to 0$, then $\mathbf{V}_{\nu,p}(h_n) = \mathbb{V}[\hat{\tau}_{\nu,p}(h_n) \mid \mathcal{X}_n] = \mathcal{V}_{+,v,p}(h_n) + \mathcal{V}_{-,\nu,p}(h_n)$, where

$$\begin{aligned}
\mathcal{V}_{+,\nu,p}(h_n) &= \frac{1}{nh_n^{2\nu}}\nu!^2 e_\nu' \Gamma_{+,p}^{-1}(h_n)\,\Psi_{+,p}(h_n)\,\Gamma_{+,p}^{-1}(h_n)\,e_\nu \\
&= \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_+^2}{f}\nu!^2 e_\nu' \Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu\,\{1+o_p(1)\} \\
\mathcal{V}_{-,v,p}(h_n) &= \frac{1}{nh_n^{2\nu}}\nu!^2 e_\nu' \Gamma_{-,p}^{-1}(h_n)\,\Psi_{-,p}(h_n)\,\Gamma_{-,p}^{-1}(h_n)\,e_\nu \\
&= \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_-^2}{f}\nu!^2 e_\nu' \Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu\,\{1+o_p(1)\}
\end{aligned}$$

(D) If $nh_n^{2p+5} \to 0$, then

$$\frac{\hat{\tau}_{\nu,p}(h_n) - \tau_\nu - h_n^{p+1-\nu}\mathbf{B}_{\nu,p,p+1}(h_n)}{\sqrt{\mathbf{V}_{\nu,p}(h_n)}} \to_d \mathcal{N}(0,1)$$

Proof. Part (B) follows immediately from Lemma B.2(B), its analogue for the left-side estimator $\left(s! e_s' \hat{\beta}_{-,\ell}(h_n)\right)$, and the linearity of conditional expectations. Part (V) also follows immediately from Lemma B.2(V), its analogue for the left-side estimator $\left(s! e_s' \hat{\beta}_{-,\ell}(h_n)\right)$, and the conditional independence of observations at either side of the threshold $(x = 0)$. Finally, part (D) follows by the same argument given in the proof of Lemma B.2(D), but now applied to the estimator $\hat{\tau}_{\nu,p}(h_n) = \Delta\hat{\mu}_{+,p}^{(\nu)}(h_n) - \Delta\hat{\mu}_{-,p}^{(\nu)}(h_n) = \nu! e_\nu' \hat{\beta}_{+,p}(h_n) - \nu! e_\nu' \hat{\beta}_{-,p}(h_n)$. This completes the proof. □

Appendix C Proofs of main results

We provide the proofs of our results in this appendix.

Proof of Claim 1. Given Assumptions 1 and 2, we can prove that $\tau^{DiDC} = \tau_1^{RDD} - \tau_0^{RDD} = \Delta Y^+ - \Delta Y^-$.

A) **Difference of RDDs.** From 1 and 2 we can derive two RDDs estimands, one for $t = 0$ and one for $t = 1$.

$$\begin{aligned} \tau_1^{RDD} &= Y_1^+ - Y_1^- \\ &= E\left[Y_{i,1}(1,1) - Y_{i,1}(0,1) - Y_{i,1}(1,0) + Y_{i,1}(0,0) \,|Z_i = z_0\right]\left(D_0^+ D_1^+ - D_0^- D_1^-\right) + \\ &\quad E\left[Y_{i,1}(0,1) - Y_{i,1}(0,0) \,|Z_i = z_0\right]\left(D_1^+ - D_1^-\right) + \\ &\quad E\left[Y_{i,0}(1,0) - Y_{i,1}(0,0) \,|Z_i = z_0\right]\left(D_0^+ - D_0^-\right) \end{aligned} \tag{C.1}$$

$$\begin{aligned} \tau_0^{RDD} &= Y_0^+ - Y0^- \\ &= E\left[Y_{i,0}(1,0) - Y_{i,0}(0,0) \,|Z_i = z_0\right]\left(D_0^+ - D_0^-\right) \end{aligned} \tag{C.2}$$

where $X^+ = \lim_{\epsilon \to 0} E(X_i|Z_i = z_0 + \epsilon)$ and $X^- = \lim_{\epsilon \to 0} E(X_i|Z_i = z_0 - \epsilon)$. By taking the difference between the equations C.1 and C.2 and under assumptions 1-4, we arrive at the differences-in-discontinuity estimand:

$$\begin{aligned} \tau^{DiDC} &= \tau_1^{RDD} - \tau_0^{RDD} \\ &= E\left[(Y_{i,1}(1,1) - Y_{i,1}(0,1))(Y_{i,1}(1,0) - Y_{i,1}(0,0))\,|Z_i = z_0\right]\left(D_0^+ D_1^+ - D_0^- D_1^-\right) + \\ &\quad E\left[Y_{i,1}(0,1) - Y_{i,1}(0,0)|Z_i = z_0\right]\left(D_1^+ - D_1^-\right) + \\ &\quad E\left[(Y_{i,1}(1,0) - Y_{i,1}(0,0))(Y_{i,0}(1,0) - Y_{i,0}(0,0))\,|Z_i = z_0\right]\left(D_0^+ - D_0^-\right) \end{aligned}$$

B) **RDD of the differences.** Using equation3 and taking the difference of the limits above and below the threshold:

$$\begin{aligned} \tau^{DiDC} &= \Delta Y^+ - \Delta Y^- \\ &= E\left[(Y_{i,1}(1,1) - Y_{i,1}(0,1))(Y_{i,1}(1,0) - Y_{i,1}(0,0))\,|Z_i = z_0\right]\left(D_0^+ D_1^+ - D_0^- D_1^-\right) + \\ &\quad E\left[Y_{i,1}(0,1) - Y_{i,1}(0,0)|Z_i = z_0\right]\left(D_1^+ - D_1^-\right) + \\ &\quad E\left[(Y_{i,1}(1,0) - Y_{i,1}(0,0))(Y_{i,0}(1,0) - Y_{i,0}(0,0))\,|Z_i = z_0\right]\left(D_0^+ - D_0^-\right) \end{aligned}$$

where $\Delta X^+ = \lim_{\epsilon \to 0} E(\Delta X_i|Z_i = z_0 + \epsilon)$ and $\Delta X^- = \lim_{\epsilon \to 0} E(\Delta X_i|Z_i = z_0 - \epsilon)$.

Therefore $\tau^{DiDC} = \tau_1^{RDD} - \tau_0^{RDD} = \Delta Y^+ - \Delta Y^-$. □

Proof. Proof of Lemma 1

Recall the definition $MSE_{\nu,p,s}(h_n) = E\left[(\hat{\tau}_{\nu,p,s}(h_n) - \tau_{\nu,p})^2\,|\chi_n\right]$. We can rewrite as

$$MSE_{\nu,p,s}(h_n) = V\left[\hat{\tau}_{\nu,p,s}(h_n)|\chi_n\right] + \left(E\left[\hat{\tau}_{\nu,p,s}(h_n) - \tau_{\nu,p}|\chi_n\right]\right)^2$$

Then, from Lemma B.4,

$$V\left[\hat{\tau}_{\nu,p,s}(h_n)|\chi_n\right] = \frac{\nu!}{nh_n^{1+2\nu}} \frac{\sigma_+^2 - \sigma_-^2}{f} e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1}\{1 + o_p(1)\}$$

and

$$E\left[\hat{\tau}_{\nu,p,s}(h_n) - \tau_{\nu,p}|\chi_n\right] = h_n^{1+p-\nu} \frac{\Delta\mu_+^{(p+1)} - (-1)^{\nu+p+s}\Delta\mu_-^{(p+1)}}{(p+1)!} \nu! e_\nu' \Gamma_p^{-1} \varphi_{p,p+1}\{1 + o_p(1)\} \tag{C.3}$$

so we can rewrite

$$MSE_{\nu,p,s}(h_n) = \frac{\mathbb{V}_{\nu,\imath}}{nh_n^{1+2\nu}}\{1 + o_p(1)\} + \left(h_n^{1+p-\nu}\mathbb{B}_{\nu,\imath,\imath+\not\Bbbk,\sim}\{1 + o_p(1)\}\right)^2$$

where $\mathbb{V}_{\nu,p} = \nu! \frac{\sigma_+^2 - \sigma_-^2}{f} e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1}$ and $\mathbb{B}_{\nu,p,p+1,s} = \frac{\Delta\mu_+^{(p+1)} - (-1)^{\nu+p+s} \Delta\mu_-^{(p+1)}}{(p+1)!} \nu! e_\nu' \Gamma_p^{-1} \varphi_{p,p+1}$. The MSE-optimal bandwidth is

$$
\begin{aligned}
{h_n^{MSE}}_{\nu,p} &= arg \min_{h_n} MSE_{\nu,p,s}(h_n) \\
&= \left(\frac{(1+2\nu)\,\mathbf{V}_{\nu,p}}{n2\,(1+p-\nu)\,\mathbf{B}_{\nu,p,p+1,s}^2} \right)^{\frac{1}{2p+3}}
\end{aligned}
$$

□

Proof. Proof of Claim 2

We define higher-order derivatives notation of the unknown regression functions as:

$$
\begin{aligned}
\Delta\mu_+^{(\nu)}(z) &= \frac{d^\nu \Delta\mu_+(z)}{dz^\nu}, & \Delta\mu_-^{(\nu)}(z) &= \frac{d^\nu \Delta\mu_(z)}{dz^\nu} \\
\Delta\mu_+^{(\nu)} &= \lim_{x\to 0^+} \Delta\mu_+^{(\nu)}(z), & \Delta\mu_-^{(\nu)} &= \lim_{x\to 0^-} \Delta\mu_-^{(\nu)}(z) \\
\Delta\mu_+ &= \lim_{x\to 0^+} \Delta\mu(z), & \Delta\mu_- &= \lim_{x\to 0^=} \Delta\mu(z)
\end{aligned}
$$

where $\Delta\mu(z) = E\left(\Delta Y_i | Z_i = z\right) = E\left(Y_{i,1} - Y_{i,0} | Z_i = z\right)$.

The notation for the RDD at time t where $\mu_t(z) = E\left(Y_{i,t} | Z_i = z\right)$ is

$$
\begin{aligned}
\mu_{t,+}^{(\nu)}(z) &= \frac{d^\nu \mu_{t,+}(z)}{dz^\nu}, & \mu_{t,-}^{(\nu)}(z) &= \frac{d^\nu \mu_(z)}{dz^\nu} \\
\mu_{t,+}^{(\nu)} &= \lim_{x\to 0^+} \mu_{t,+}^{(\nu)}(z), & \mu_{t,-}^{(\nu)} &= \lim_{x\to 0^-} \mu_{t,-}^{(\nu)}(z) \\
\mu_{t,+} &= \lim_{x\to 0^+} \mu(z), & \mu_{t,-} &= \lim_{x\to 0^=} \mu(z)
\end{aligned}
$$

Then we derive the following:

$$
\Delta\mu_+^{(\nu)} = \mu_{1,+}^{(\nu)} - \mu_{0,+}^{(\nu)} \qquad \Delta\mu_-^{(\nu)} = \mu_{1,}^{(\nu)} - \mu_{0,+}^{(\nu)} \tag{C.4}
$$

It follows that $B\left[\hat{\tau}_{\nu,p}^{DiDC}(h_n)\right] = B\left[\hat{\tau}_{1,\nu,p}^{RDD}(h_n)\right] - B\left[\hat{\tau}_{0,\nu,p}^{RDD}(h_n)\right]$ □

Proof. Proof of Calim 3

$$
\begin{aligned}
\Delta\mu(z) =& E\left[\Delta Y_i | Z_i = z\right] \\
=& E\left[Y_{i,1} - Y_{i,0} | Z_i = z\right] \\
=& E\left[Y_{i,1} | Z_i = z\right] - E\left[Y_{i,0} | Z_i = z\right] \\
=& \mu_1(z) - \mu_0(z)
\end{aligned}
$$

Therefore,

$$
\Delta\mu_+^{(\nu)} = \mu_{+,1}^{(\nu)} - \mu_{+,0}^{(\nu)}
$$

Then, if the shapes of the DGPs are time-invariant for each side of the threshold, then $\mu_{+,1}^{(\nu)} - \mu_{+,0}^{(\nu)}$ and $\mu_{-,1}^{(\nu)} - \mu_{-,0}^{(\nu)}$, leading to $\Delta\mu_+^{(\nu)} = \Delta\mu_-^{(\nu)} = 0$ □

Appendix D Simulations

D.1 Bandwidth for Testing if Confounding Effects are Constant

This appendix details into the critical issue of bandwidth selection the Wald test to test Assumption3. As mentioned before, this assumption is vital for the difference-in-discontinuitis method, as the non validity means all estimations would be unvalid. The procedure for testing involves considering the following stacked RDDs regression model:

$$Y_i = \alpha_i + \sum_{k=1}^{K} \left[\beta_{-k} \left(Z_i - z_0\right) + \theta_{-k} D \left(Z_i - z_0\right)\right] T_{-k} + \varepsilon_i \tag{8}$$

where T_{-k} is a dummy indicating which period the data belongs to, that is, the RDD of which period it's running. The null hypothesis for the Wald test is:

$$H_0: \quad \theta_0 = \theta_{-1} = ... = \theta_{-K}$$

One central question arises: how should we determine the optimal bandwidth for Equation 8. In this section, we outline the methodology used for bandwidth selection and present simulation results in two different scenarios: one with identical Data Generating Processes for pre- and post-treatment periods and another with different DGPs, both while considering confounding factors. The first scenario aligns with the framework presented in Model 1, as detailed in Appendix D.2.1 while the second encompasses both models 1 and 3 (Appendix D.2.1).

To select bandwidths, we employ the `rdrobust` package and estimate bandwidths separately for each RDD. Next, we cross the bandwidths and conduct RDD regressions with each of them, evaluating the bias and root-mean-squared errors (RMSE) of the estimated parameters. The chosen bandwidth are the one that minimizes RMSE, alongside the largest and smallest bandwidths, and the CCT (Calonico et al., 2014) bandwidth for all observations regardless of time dummies.

The various bandwidths selected for produced similar results. This finding suggests that the choice of bandwidth does not substantially affect the accuracy of the test. It is essential to note that while this results says this, the choice of bandwidth for the DiDC estimation of the treatment effect remains a crucial consideration of the method.

D.2 Data Generating Processes

We provide further details on each of the data generating processes (DGPs) employed in our simulation studies. Both GDPs employ the same simulation setup based on model 3 from Calonico et al. (2014), with the key difference being the presence of time-fixed confounding factors at the threshold in one of the settings. We perform 1000 replications for each simulation and for each replication, we use a sample size of $n = 500$, with

$$Y_{i,t} = \mu_{it}(Z_i) + \varepsilon_{i,t}, \quad Z_i \sim (2\mathcal{B}(2,4) - 1),$$
$$\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$$

where $\mathcal{B}(p_1, p_2)$ is a beta distribution with parameters p_1 and p_2, $\sigma_\varepsilon = 0.1295$.

D.2.1 Model 1: Identical DGPs with time-invariant confounders at the threshold

This model includes a time-invariant confounder, **c**, to represent pre-existing discontinuities at the threshold that affect the outcome Y_i. Here, units with $Z_i \geq 0$ receive a treatment **c** and units with $Z_i < 0$ do not. At the same time, units with $Z_i \geq 0$ have a different data generating function than units with $Z_i < 0$. The data generating functions $\mu_0(z)$ and $\mu_1(z)$, for periods $t = 0$ and $t = 1$ respectively are as follows

$$\mu_0(z) = \begin{cases} 0.48 + 1.27z - 0.5 \cdot 7.18z^2 + 0.7 \cdot 20.21z^3 + 1.1 \cdot 21.54z^4 + 1.5 \cdot 7.33z^5 & \text{if } z < 0 \\ 0.52 + 0.84z - 0.1 \cdot 3.00z^2 - 0.3 \cdot 7.99z^3 - 0.1 \cdot 9.01z^4 + 3.56z^5 + \mathbf{c} & \text{if } z \geq 0 \end{cases}$$

$$\mu_1(z) = \begin{cases} 0.48 + 1.27z - 0.5 \cdot 7.18z^2 + 0.7 \cdot 20.21z^3 + 1.1 \cdot 21.54z^4 + 1.5 \cdot 7.33z^5 & \text{if } z < 0 \\ 0.52 + 0.84z - 0.1 \cdot 3.00z^2 - 0.3 \cdot 7.99z^3 - 0.1 \cdot 9.01z^4 + 3.56z^5 + \mathbf{c} + \tau & \text{if } z \geq 0 \end{cases}$$

D.2.2 Model 2: Identical DGPs

This model follows precisely model 3 from Calonico et al. (2014) for both periods, the only difference between them being that in period $t = 1$ there is the effect of the treatment τ for units with $Z - i \geq 0$. No other discontinuities that could affect the outcome exist at the threshold $z_0 = 0$. The data generating functions $\mu_0(z)$ and $\mu_1(z)$ are as follows

$$\mu_0(z) = \begin{cases} 0.48 + 1.27z - 0.5 \cdot 7.18z^2 + 0.7 \cdot 20.21z^3 + 1.1 \cdot 21.54z^4 + 1.5 \cdot 7.33z^5 & \text{if } z < 0 \\ 0.52 + 0.84z - 0.1 \cdot 3.00z^2 - 0.3 \cdot 7.99z^3 - 0.1 \cdot 9.01z^4 + 3.56z^5 & \text{if } z \geq 0 \end{cases}$$

$$\mu_1(z) = \begin{cases} 0.48 + 1.27z - 0.5 \cdot 7.18z^2 + 0.7 \cdot 20.21z^3 + 1.1 \cdot 21.54z^4 + 1.5 \cdot 7.33z^5 & \text{if } z < 0 \\ 0.52 + 0.84z - 0.1 \cdot 3.00z^2 - 0.3 \cdot 7.99z^3 - 0.1 \cdot 9.01z^4 + 3.56z^5 + \tau & \text{if } z \geq 0 \end{cases}$$

D.2.3 Model 3: Different DGPs with time-invariant confounders at the threshold

This model follows model 3 from Calonico et al. (2014) and includes a time-invariant confounder, **c**, to represent pre-existing discontinuities at the threshold that affect the outcome Y_i at $t = 0$, and a linear model for time $t = 1$. The data generating functions $\mu_0(z)$ and $\mu_1(z)$, for periods $t = 0$ and $t = 1$ respectively are as follows

$$\mu_0(z) = \begin{cases} 0.48 + 1.27z - 0.5 \cdot 7.18z^2 + 0.7 \cdot 20.21z^3 + 1.1 \cdot 21.54z^4 + 1.5 \cdot 7.33z^5 & \text{if } z < 0 \\ 0.52 + 0.84z - 0.1 \cdot 3.00z^2 - 0.3 \cdot 7.99z^3 - 0.1 \cdot 9.01z^4 + 3.56z^5 + \mathbf{c} & \text{if } z \geq 0 \end{cases}$$

$$\mu_1(z) = \begin{cases} 0.48 + 1.4z & \text{if } z < 0 \\ 0.52 + 0.1z + \mathbf{c} + \tau & \text{if } z \geq 0 \end{cases}$$

D.2.4 Model 4: Different DGPs

Here the model for $t = 0$ follows precisely model 3 from Calonico et al. (2014) and the model for $t = 1$ is linear. No discontinuities that could affect the outcome exist at the threshold $z_0 = 0$ other than the treatment effect τ introduced for $t = 1$. The data generating functions $\mu_0(z)$ and $\mu_1(z)$ are as follows

$$\mu_0(z) = \begin{cases} 0.48 + 1.27z - 0.5 \cdot 7.18z^2 + 0.7 \cdot 20.21z^3 + 1.1 \cdot 21.54z^4 + 1.5 \cdot 7.33z^5 & \text{if } z < 0 \\ 0.52 + 0.84z - 0.1 \cdot 3.00z^2 - 0.3 \cdot 7.99z^3 - 0.1 \cdot 9.01z^4 + 3.56z^5 + \mathbf{c} & \text{if } z \geq 0 \end{cases}$$

$$\mu_1(z) = \begin{cases} 0.48 + 1.4z & \text{if } z < 0 \\ 0.52 + 0.1z + \tau & \text{if } z \geq 0 \end{cases}$$

Appendix E Empirical illustration

We replicate the analysis conducted by Grembi et al. (2016) once more, this time using the bandwidth selection method proposed by Ludwig and Miller (2007) as she did. The use of these alternative bandwidths does not significantly alter the estimated treatment effects; they remain similar in magnitude.

Table E.1: Effects of relaxing a fiscal rule - Ludwig & Miller (2007) Bandwidths

Estimators	Deficit	Fiscal Gap	Taxes
Diif-in-Disc (Grembi et al, 2016)	9.454	48.469	-34.748
	(4.343)	(23.315)	(20.166)
Bandwidth	1498	833	684
Observations	5858	3438	1536
Diff-in-Disc (RDD of Δ)	16.979	35.302	-16.958
	(8.859)	(12.449)	(7.473)
Bandwidth	887	908	1047
Observations	1379	1278	1254
Diff-in-Disc (Δ of RDDs)	17.895	39.585	-27.620
	(4.713)	(12.729)	(10.761)
Bandwidth	961	647	695
Observations	1163	1156	1246